A dimension reduction approach for conditional Kaplan-Meier estimators

Weiyu Li* and Valentin Patilea†

January 2015

Abstract

In many situations, the quantities of interest in survival analysis are smooth, closed-form expression functionals of the law of the observations. In such cases, one can easily derive nonparametric estimators for the quantities of interest by plugging-into the functional the nonparametric estimators of the law of the observations. However, in the presence of multivariate covariates, the nonparametric estimation suffers from the curse of dimensionality. In this paper, a new general dimension reduction approach for conditional models in survival analysis is proposed. It is based on semiparametric index modeling of the law of the observations.

Keywords. Bootstrap, Cure models, Kernel smoothing, Semiparametric regression, Single-index, $U-$statistics

I Introduction

Survival data, also called duration or time to event data, are often incomplete due to the presence of censoring. In such cases, the model has to take into account the censoring mechanism in order to avoid substantial biases. Several types of models and inference approaches have been developed, among which the proportional hazards model and the partial likelihood estimation method of Cox (1972) with right-censored data are perhaps the most famous. In a more flexible, nonparametric regression perspective, the conditional distribution of the response given the covariates in the presence of right censoring is usually estimated by the Beran (1981)'s estimator, also called the conditional Kaplan-Meier estimator. See also Dabrowska (1989, 1992) who pointed out that Beran's estimator is a smooth functional of the kernel regression estimator of the conditional law of the observations. An obvious drawback of the nonparametric approach is the curse of dimensionality that one faces with multi-dimensional covariates.

*CREST (Ensai) & IRMAR (UEB), France; weiyu.li@ensai.fr

†Corresponding author. CREST (Ensai) & IRMAR (UEB), France; patilea@ensai.fr

A common remedy to the curse of dimensionality consists in considering a dimension reduction device before applying the nonparametric smoother. Semiparametric index regression is a common example of dimension reduction approach that has been already used to estimate conditional laws with complete data, see for instance Hall & Yao (2005), Chiang & Huang (2012), Ma & Zhu (2013) and Lee *et al.* (2013). Index regression is also the idea we follow in this paper where the responses could be censored. Semiparametric index estimation of a conditional law in the presence of right censoring was considered recently by Xia *et al.* (2010), Bouziz & Lopez (2010) and Strzalkowska-Kominiak & Cao (2013). However, the existing approaches impose the dimension reduction directly on the conditional law of the lifetime of interest. In general, this results in rather complicated procedures.

In this paper we show that there is an alternative, easier way to introduce the model assumptions. For this we recall Dabrowska's remark that in many situations, the quantities of interest in survival analysis are smooth, closed-form expression functionals of the law of the observations. This is, for instance, the case for the conditional law of the lifetime of interest under random right censoring, but also for other quantites, as for instance the conditional probability of being cured in cure survival models. Keeping this in mind, we propose to impose the dimension reduction index hypothesis directly on the conditional law of the observed variables, there are typically the possibly censored lifetime and the indicator for the presence of censoring. Next, we apply the smooth functionals to the estimator of the conditional law of the observations. This results in semiparametric estimators of the quantites of interest that avoid the curse of dimensionality. The new methodology allows to test the dimension reduction assumption and extends to other dimension reduction methods. Moreover, it can be applied to more general censoring mechanisms and closed-form expression functionals.

The paper is organized as follows. In section II we recall the general framework of right-censored data. In section III we introduce the new semiparametric index-regression approach and we focus on the single-index case. We also reconsider the general construction of Kaplan-Meier functionals in the context of the new dimension reduction idea. We end section III with a discussion on the existing single-index approaches. In particular, we shed new light on these approaches and show that our framework is not more restrictive. In section IV we propose a general semiparametric estimation method for the index and we prove the $\sqrt{n}-$normality of the index estimator. The asymptotic result is derived under mild conditions, in particular the covariates need not to be bounded or absolutely continuous, and no trimming is required. Since the estimation procedure is designed in the space of the observations, the $\sqrt{n}-$normality result remains valid even when the usual identification assumption used in the survival analysis with covariates, that is the censoring is noninformative given the covariates, fails. We end section section IV with some proposals for building confidence intervals for the index vector and for testing the dimension reduction assumption against general nonparametric alternatives. In section V we reconsider the estimation of the conditional law of the lifetime of interest using the Beran estimator with the estimated index. We derive an i.i.d. representation of the new single-index estimator. As a consequence of this representation, we introduce and prove the asymptotic normality of a new single-index estimator of the cure fraction. Our new methodology is illustrated by empirical experiments using simulated and real data. In particular, we provide a new point of view for modeling Stanford heart transplantation data. We end the paper with discussions on possible extensions of the new approach and an example of a more complicated censoring mechanism that generates a closed form expression map between the observations and the conditional law of the lifetime of interest. The assumptions and the main proofs are postponed to the Appendix. A Supplementary Material completes our work with some additional technical results.

II The framework

Let T denote the lifetime of interest that takes values in $(-\infty, \infty]$. Consider the situation where one observes independent copies of Y, δ and X, where Y is a real-valued random variable, δ is an indicator variable and X is a covariate taking values in some space $\mathcal{X}$. For the moment, we do not need to make any specific assumption about $\mathcal{X}$. The indicator variable reveals whether Y is precisely the lifetime of interest, or Y is only a random quantity smaller than T. In other words,

$$\delta = 1 \quad \text{if} \quad Y = T \qquad \text{and} \qquad \delta = 0 \quad \text{if} \quad Y < T.$$

The purpose is to estimate the law of T given X. The conditional probability of the event $\{T = \infty\}$ will be allowed to be positive.

The observations are characterized by the conditional sub-probabilities

$$\begin{aligned} H_1((-\infty, t] \mid x) &= \mathbb{P}(Y \leq t, \delta = 1 \mid X = x) \\ H_0((-\infty, t] \mid x) &= \mathbb{P}(Y \leq t, \delta = 0 \mid X = x), \qquad t \in \mathbb{R}, \ x \in \mathcal{X}. \end{aligned}$$

Then the law of Y is characterized by

$$H((-\infty, t] \mid x) = \mathbb{P}(Y \leq t \mid X = x) = H_0((-\infty, t] \mid x) + H_1((-\infty, t] \mid x).$$

We suppose Y is finite, that means

$$H((-\infty, \infty) \mid x) = 1, \qquad \forall x \in \mathcal{X}.$$

The usual way to model this situation in order to estimate the conditional law T is to consider that there exists a random variable C, the right-censoring time, and

$$Y = T \wedge C, \qquad \delta = \mathbf{1}\{T \leq C\}.$$

Using suitable identification conditions, the conditional law of T given X could be expressed as a closed-form expression functional of $H_0(\cdot \mid x)$ and $H_1(\cdot \mid x)$ and thus could be easily estimated by plugging in nonparametric estimates of $H_0(\cdot \mid x)$ and $H_1(\cdot \mid x)$. Such an estimator of the conditional law of T given X is usually called a conditional

Kaplan-Meier estimator. See Beran (1981), Dabrowska (1989), Van Keilegom & Veraverbeke (1996). All these approaches suffer from the curse of dimensionality when $\mathcal{X}$ is of higher dimension than the real line. Here, we propose a dimension reduction approach for estimating $H_0(\cdot \mid x)$ and $H_1(\cdot \mid x)$ and we will focus on the single-index case. This will result in a natural single-index estimator of the conditional law of T given X.

III Single-index modeling under random censoring

In this section, we introduce our general dimension reduction approach using the law of the observations and we show how it induces a dimension reduction for the quantities of interest. For the sake of simplicity, we focus on the single-index case, though the idea is clearly more general. In particular, we shed new light on the existing conditional distribution single-index models.

III.1 Dimension reduction for modeling the conditional law of the observations

Let us consider that the conditional law of (Y, δ) satisfies a dimension reduction condition. More precisely, let $\mathcal{B}$ be a parameter space, $q \geq 1$ and let

$$\lambda : \mathcal{X} \times \mathcal{B} \to \mathbb{R}^q$$

be a given map such that

$$(Y, \delta) \perp X \mid \lambda(X, \beta_0).$$

for some $\beta_0 \in \mathcal{B}$. Typically, the value of q is much smaller than the dimension of the covariates space $\mathcal{X}$. Although the dimension reduction approach is potentially more general, we will focus on the case

$$\mathcal{X} = \mathbb{R}^p, \quad \mathcal{B} \subset \mathbb{R}^p \quad \text{and} \quad \lambda(X, \beta) = X^\top \beta \in \mathbb{R}.$$

(Here and in the following, a vector is a column matrix and for any matrix A, $A^\top$ denotes its transpose.) In other words, we focus on the single-index modeling of the conditional law of the observations (Y, δ) given finite dimension covariates. Hence we suppose

$$(Y, \delta) \perp X \mid X^\top \beta_0 \tag{III.1}$$

for some $\beta_0 \in \mathcal{B}$. Similar modeling was considered, for instance, by Delecroix *et al.* (2003), Hall & Yao (2005), Chiang & Huang (2012), Ma & Zhu (2013) and Lee *et al.* (2013). However, none of the previous approaches seems appropriate in the present framework where (Y, δ) does not have a conditional density and X is not necessarily required to be absolutely continuous.

We can rewrite the condition (III.1) using the subdistributions H_0 and H_1. For any $d \in \{0, 1\}$ and $t \in \mathbb{R}$, let

$$H_{d,\beta}((-\infty, t] \mid z) = \mathbb{P}(Y \leq t, \delta = d \mid X^\top \beta = z), \qquad z \in \mathbb{R},$$

and $H_{d,\beta}(dt \mid z)$ be the associated measures. Then, condition (III.1) means there exists $\beta_0 \in \mathcal{B}$ such that

$$H_d((-\infty,t] \,|\, X) = H_{d,\beta_0}((-\infty,t] \,|\, X^\top\beta_0), \;\; \text{almost surely (a.s.), for any } d \in \{0,1\}, \; t \in \mathbb{R}.$$

In general, the parameter value of β_0 is unknown and has to be estimated. For this purpose, in section IV.1, we propose a new estimation method to derive a $\sqrt{n}-$asymptotically normal semiparametric estimator of β_0. The covariates could be unbounded and no trimming is required. For the moment, let us consider that β_0 is given.

III.2 Conditional single-index Kaplan-Meier functionals

If C is a random right-censoring variable and $Y = T \wedge C, \;\; \delta = \mathbf{1}\{T \le C\}$, one can link the conditional laws of C and T to the conditional subdistributions of the observations. More precisely, for any $x \in \mathcal{X}$ and $t \in \mathbb{R}$, we have

$$\begin{aligned} H_1((-\infty,t] \mid x) &= \int_{(-\infty,t]} \mathbb{P}(C \ge T \mid X = x, T = s) F_T(ds \mid x), \\ H_0((-\infty,t] \mid x) &= \int_{(-\infty,t]} \mathbb{P}(T > C \mid X = x, C = s) F_C(ds \mid x), \end{aligned} \qquad \text{(III.2)}$$

where $F_T(dt \mid x)$ and $F_C(dt \mid x)$ are the conditional distributions functions of T and C given that $X = x$. To be able to build consistent estimates of the quantities of interest from the data, some identification assumptions are required for any value x. For this purpose one could use the usual conditional independence assumptions imposed in survival analysis

$$T \perp C \mid X. \qquad \text{(III.3)}$$

For the sake of simplicity, in the following we also assume $\mathbb{P}(C = T) = 0$. Then the equations (III.2) become

$$\begin{aligned} H_1(dt \mid x) &= F_C([t,\infty) \mid x) F_T(dt \mid x) \\ H_0(dt \mid x) &= F_T([t,\infty] \mid x) F_C(dt \mid x). \end{aligned}$$

Following our dimension reduction approach, we assume

$$(T,C) \perp X \mid X^\top\beta_0 \qquad \text{(III.4)}$$

for some vector $\beta_0 \in \mathcal{B}$. Our model assumptions imply that β_0 is also the vector satisfying the condition (III.1). Moreover, the equations (III.2) could be rewritten under the form

$$\begin{aligned} H_{1,\beta}(dt \mid z) &= F_{C,\beta}([t,\infty) \mid z) F_{T,\beta}(dt \mid z) \\ H_{0,\beta}(dt \mid z) &= F_{T,\beta}([t,\infty] \mid z) F_{C,\beta}(dt \mid z) \end{aligned} \qquad \text{(III.5)}$$

with $\beta = \beta_0$, where for any $t \in \mathbb{R}$ and $z \in \mathbb{R}$,

$$F_{T,\beta}((-\infty,t] \mid z) = \mathbb{P}(T \le t \mid X^\top\beta = z), \quad F_{C,\beta}((-\infty,t] \mid z) = \mathbb{P}(C \le t \mid X^\top\beta = z),$$

and $F_{T,\beta}(dt \mid z)$ and $F_{C,\beta}(dt \mid z)$ are the associated measures.

It is well known that, for any fixed β, the equations (III.5) could be explicitly solved for $F_{T,\beta}$ and $F_{C,\beta}$. Indeed, let us consider the conditional cumulative hazard measures

$$\Lambda_{T,\beta}(dt \mid z) = \frac{F_{T,\beta}(dt \mid z)}{F_{T,\beta}([t,\infty] \mid z)} \quad \text{and} \quad \Lambda_{C,\beta}(dt \mid z) = \frac{F_{C,\beta}(dt \mid z)}{F_{C,\beta}([t,\infty) \mid z)}, \qquad z \in \mathbb{R}.$$

Then the model equations (III.5) could be solved for any z and t such that $H_\beta([t,\infty) \mid z) > 0$ and this yields

$$\Lambda_{T,\beta}(dt \mid z) = \frac{H_{1,\beta}(dt \mid z)}{H_\beta([t,\infty) \mid z)} \quad \text{and} \quad \Lambda_{C,\beta}(dt \mid z) = \frac{H_{0,\beta}(dt \mid z)}{H_\beta([t,\infty) \mid z)},$$

where $H_\beta(\cdot \mid z) = H_{0,\beta}(\cdot \mid z) + H_{1,\beta}(\cdot \mid z)$. Then, we could define the *single-index Kaplan-Meier functionals*

$$\begin{aligned} F_{T,\beta}((t,\infty] \mid z) &= \prod_{-\infty<s\leq t} \{1 - \Lambda_{T,\beta}(ds \mid z)\}, \\ F_{C,\beta}((t,\infty) \mid z) &= \prod_{-\infty<s\leq t} \{1 - \Lambda_{C,\beta}(ds \mid z)\}, \qquad t \in \mathbb{R}, \end{aligned} \tag{III.6}$$

where $\prod_{s\in A}$ stands for the product-integral over the set A (see Gill & Johansen, 1990). Let us remember that although Y takes finite values on the real line, we allow for a positive probability for the event $\{T = \infty\}$. Hence, implicitly we also assume that the conditional subdistributions H_0 and H_1 are such that $F_{C,\beta}((-\infty,\infty) \mid z) = 1$, for any z. This mild condition is satisfied only if, for each value of the covariate, the upper bound of support of $H_1(\cdot \mid x)$, that could be finite or infinite at this stage, is smaller or equal to the upper bound of the support of $H_0(\cdot \mid x)$.

Let us point out that, for each fixed vector β, *any* two conditional sub-probabilities H_0 and H_1 satisfying the mild condition $F_{C,\beta}((-\infty,\infty) \mid z) = 1$, for any z, define uniquely the functionals $F_{T,\beta}(\cdot \mid z)$ $F_{C,\beta}(\cdot \mid z)$. If the assumptions (III.4) and (III.3) hold true, then $F_{T,\beta_0}(\cdot \mid z)$ is precisely the single-index conditional probability distribution of T. In particular, this remark explains why in the following asymptotic results it will not be needed to impose the assumptions (III.4) and (III.3) and only the single-index condition (III.1) will be required.

Semiparametric estimators for $\Lambda_{T,\beta_0}(\cdot \mid x^\top\beta_0)$ and $F_{T,\beta_0}(\cdot \mid x^\top\beta_0)$ are easily built by plugging-into the previous formulae an estimator of β_0 and a nonparametric estimator of the conditional subdistributions $H_{d,\beta}(\cdot \mid x^\top\beta)$, $d \in \{0,1\}$. This idea will be detailed in the following sections. For now, let us cast new light on some recent single-index approaches in survival analysis literature.

III.3 Comparison with the previous approaches

Conditional distribution single-index models under random right censoring were recently proposed by Bouaziz & Lopez (2010) and Strzalkowska-Kominiak & Cao (2013). See also

Lu (2010) and Strzalkowska-Kominiak & Cao (2014). These contributions impose the existence of a conditional density for Y and use a likelihood criterion to estimate the index β_0. Such likelihood approaches require quite involved technical assumptions. Our approach is simpler and could provide interesting insight in their identification assumptions, as is shown in the following.

A possible way to solve the system of equations (III.2) for F_T is to follow Stute (1993) and suppose that

$$T \perp C \qquad \text{and} \qquad \mathbb{P}(C \geq T \mid X, T) = \mathbb{P}(C \geq T \mid T). \tag{III.7}$$

Then the equation (III.2) becomes

$$H_1(dt \mid x) = F_C([t, \infty))F_T(dt \mid x),$$

where $F_C(\cdot)$ is the marginal distribution function of C. Basically, the likelihood of Bouaziz & Lopez (2010) involves only this equation. To implement their estimation method, Bouaziz & Lopez (2010) replaced $F_C(\cdot)$ by the classical Kaplan-Meier estimator. Let us point out that a single-index assumption on $F_T(dt \mid x)$ induces a single-index structure on $H_1(dt \mid x)$, but also on $H_0(dt \mid x)$.

Strzalkowska-Kominiak & Cao (2013) also imposed the conditions (III.7) but they aimed at using a more adapted likelihood which corresponds to involving both equations (III.2) in the construction of the likelihood. Therefore they also imposed the condition $T \perp C \mid X$. Then one has the relationships

$$H_1(dt \mid x) = F_C([t, \infty))F_T(dt \mid x), \qquad H_0(dt \mid x) = F_T([t, \infty) \mid x)F_C(dt). \tag{III.8}$$

Next, the single-index condition is imposed on the density of $F_T(dt \mid x)$ and the index estimated by maximum likelihood. Again, for implementation $F_C(\cdot)$ was replaced by the Kaplan-Meier estimator. Note that the assumptions of Strzalkowska-Kominiak & Cao (2013) induce a single-index assumption on *both* $H_1(dt \mid x)$ and $H_0(dt \mid x)$.

In both approaches we discussed above, one could use the observations (Y, δ) to estimate much easily the index β_0, for instance using the estimation method we present in the next session, and then invert the equations above to recover $F_T(dt \mid x)$.

Let us also provide an interpretation of the contribution of Xia *et al.* (2010) in the case of a single-index assumption. Under the condition $T \perp C \mid X$, one has $H([t, \infty] \mid x) = F_T([t, \infty] \mid x)F_C([t, \infty] \mid x)$ so that one could rewrite the equation (III.2) under the form

$$\frac{h_1(t \mid x)dt}{H([t, \infty] \mid x)} = \frac{F_T(dt \mid x)}{F_T([t, \infty] \mid x)} = \lambda_T(t \mid x)dt,$$

where h_1 is the Radon-Nikodym derivative of H_1, that is $H_1(t \mid x) = h_1(t \mid x)dt$, and $\lambda_T(\cdot \mid x)$ is the conditional hazard function of T given $X = x$. In the case of a single-index setup, $T \perp X \mid X^\top \beta_0$ and this induces the single-index condition on the conditional hazard function, that is $\lambda_T(\cdot \mid x) = \lambda_T(\cdot \mid x^\top\beta_0)$. Next, by elementary properties of the convolution, for a kernel $\Gamma(\cdot)$ and a bandwidth b,

$$\frac{\mathbb{E}\left[\delta\Gamma((Y-t)/b)/b \mid X=x\right]}{H([t,\infty] \mid x)} \to \frac{h_1(t\mid x)dt}{H([t,\infty]\mid x)} = \lambda_T(t\mid x^\top\beta_0), \quad \text{as } b\to 0.$$

To estimate β_0, Xia *et al.* (2010) used an average-derivative approach based on the response $\delta b^{-1}\Gamma((Y-t)/b)H^{-1}([t,\infty] \mid x)$ with t running over the whole sample space of Y. In their section 3.2, the authors proposed a refined estimation by imposing the single-index condition on $H([t,\infty] \mid x)$ too; see also their assumption E2. In view of the equations (III.8) one can easily realize that, in the single-index case, our dimension reduction assumption (III.4) is practically the same as that of Xia *et al.* (2010). However, our estimation approach is simpler and could be implemented under milder assumptions. Moreover, as explained in section VII below, we could also consider multi-index assumptions and investigate other types of censoring mechanisms.

IV The index estimation method

The observations $(Y_i,\delta_i,X_i)\in\mathbb{R}\times\{0,1\}\times\mathbb{R}^p$ $1\leq i\leq n$, are independent copies of (Y,δ,X).

IV.1 Semiparametric estimation of the index

For identification purposes, for any $\beta\in\mathcal{B}$ we fix the first component of β equal to 1. Hence $\mathcal{B}\subset\{1\}\times\mathbb{R}^{p-1}$ and we suppose that $\beta_0\in\mathcal{B}$ is the unique vector satisfying condition (III.1). Let f_β be the density of $X^\top\beta$ that is supposed to exist. Let us point out that $X^\top\beta$ may have a density for any $\beta\in\mathcal{B}$ even if X is not an absolutely continuous vector of covariates. In particular, discrete covariates are allowed.

Next, let

$$U(t,d;\beta) = \left\{\mathbf{1}\{Y\leq t, \delta=d\} - H_{d,\beta}([0,t]\mid X^\top\beta)\right\} f_\beta(X^\top\beta),$$

with $t\in\mathbb{R}$, $d\in\{0,1\}$, $\beta\in\mathcal{B}$, and let $U_i(t,d;\beta)$ be the same map applied to (Y_i,δ_i,X_i). We have the following obvious equivalence

$$\mathbb{E}[U(t,d;\beta)\mid X]=0 \;\; a.s. \;\; \forall (t,d)\in\mathbb{R}\times\{0,1\} \quad \Leftrightarrow \quad (Y,\delta)\perp X\mid X^\top\beta. \qquad \text{(IV.1)}$$

Hence β_0 is the value of the parameter that makes the zero-mean conditional expectation conditions on the left-hand side of the equivalence to hold. Then, the idea is to replace that conditional expectation conditions indexed by t and k by a more convenient marginal moment condition. Let $\omega(\cdot)$ be some function defined on $\mathbb{R}^p$ and having strictly positive integrable Fourier Transform. Define

$$I(\beta) = \int_{\mathbb{R}\times\{0,1\}} \mathbb{E}\left[\omega(X_1-X_2)U_1(t,d;\beta)U_2(t,d;\beta)\right] d\mu(t,d), \qquad \text{(IV.2)}$$

where μ is some measure on $\mathbb{R} \times \{0,1\}$. Following the idea of Li & Patilea (2014), it can be shown that

$$I(\beta) \geq 0 \text{ and } I(\beta) = 0 \quad \text{if and only if} \quad \mathbb{E}[U(t,d;\beta) \mid X] = 0 \;\; a.s. \;\; \forall t \in \mathbb{R}, \; d \in \{0,1\}. \tag{IV.3}$$

The justification of the statement (IV.3) is provided in the Appendix. That means

$$I(\beta_0) = 0 \qquad \text{and} \qquad I(\beta) > 0, \;\; \forall \beta \in \mathcal{B}, \; \beta \neq \beta_0.$$

The estimation idea is to make some choice for $\omega(\cdot)$ and μ, to build a sample based approximation of $I(\beta)$ and to minimize it with respect to β.

Let us take $\omega(x) = \exp(-\|x\|^2/2)$, $x \in \mathbb{R}^p$, and μ equal to $F_{Y,\delta}$ the distribution function of the observations (Y,δ). Since $F_{Y,\delta}$ is unknown, in the applications one could approximate it by $\widehat{F}_{n,Y,\delta}$ the empirical distribution of the sample $(Y_1,\delta_1),\cdots,(Y_n,\delta_n)$. We will show that this substitution does not affect the asymptotic results. For $t \in \mathbb{R}$ and $d \in \{0,1\}$, let

$$\widehat{U}_i(t,d;\beta) = \frac{1}{n-1}\sum_{k=1}^{n} \{\mathbf{1}\{Y_i \leq t; \delta_i = d\} - \mathbf{1}\{Y_k \leq t; \delta_k = d\}\} \frac{1}{g} L_{ik}(\beta, g), \tag{IV.4}$$

where $L_{ik}(\beta,g) = L((X_i - X_k)^\top \beta / g)$ and $L(\cdot)$ is a bounded univariate kernel. Let us define the empirical approximation of $I(\beta)$ as

$$\widehat{I}_n(\beta) = \frac{1}{n^2}\sum_{1\leq i,j\leq n}\left[\frac{1}{n}\sum_{1\leq l\leq n}\widehat{U}_i(Y_l,\delta_l;\beta)\widehat{U}_j(Y_l,\delta_l;\beta)\right]\omega_{ij},$$

where $\omega_{ij} = \omega(X_i - X_j)$. Then the estimator of β is defined as

$$\widehat{\beta} = \arg\min_{\beta \in \mathcal{B}} \widehat{I}_n(\beta). \tag{IV.5}$$

Proposition 4.1 *Suppose that the condition (III.1) and Assumption VIII.1 hold true. Then* $\sqrt{n}(\widehat{\beta} - \beta_0)$ *converges in law to a centered multivariate normal distribution.*

The proof of Proposition 4.1 is relegated to the Appendix. Let us point out that in our proof X needs neither have a density nor a bounded support, and no trimming is required in the estimation procedure, as commonly imposed for estimating single-index regression models. See, for instance, Horowitz (2009). Moreover, the technical conditions only concern the observed variables (Y,δ,X) and the kernel $L(\cdot)$.

The asymptotic variance of $\widehat{\beta}$ has a complicated form presented in the proof of Proposition 4.1. A convenient bootstrap procedure designed to approximate the asymptotic law of $\widehat{\beta}$ will be proposed in the following. However, for the purpose of studying the theoretical properties of the nonparametric estimates of the conditional Kaplan-Meier functionals like $F_{T,\beta_0}(\cdot \mid z)$, one only needs to know that $\widehat{\beta}$ converges β_0 at the rate $O_{\mathbb{P}}(n^{-1/2})$.

IV.2 Implementation aspects

Dimension reduction is a convenient vehicle to go beyond the parametric modeling with covariates and use more flexible approaches in an effective way. In this section we discuss some aspects related to the implementation of our dimension reduction inference approach in survival data analysis. More precisely, we complete our estimation methods with indications how the dimension reduction assumption could be tested and confidence intervals for β_0 could be built. Before proceeding to that, let us mention that one could easily remark from the proof that the conclusion of Proposition 4.1 remains valid if ω_{ij} is replaced by $\omega_{ij}\mathbf{1}\{i \neq j\}$ in the definition of $\widehat{I}_n(\beta)$. However, our empirical experience indicates that keeping the diagonal terms in the definition of $\widehat{I}_n(\beta)$ leads to more stable numerical results. Moreover, the $\sqrt{n}-$asymptotic normality is expected to remain true even if the single-index condition (III.1) does not hold. However, in such a case β_0 has to be replaced by $\overline{\beta} = \arg\min_{\beta\in\mathcal{B}} I(\beta)$. See also Li & Patilea (2014) for a similar situation in the case without censoring.

IV.2.1 Single-index assumption check

Single-index assumptions are very common dimension reduction devices. However, one should be able to check whether such assumptions are realistic for the data at hand. Following an approach introduced by Maistre & Patilea (2014), we propose a formal test of the single-index assumption (III.1) against general alternatives.

For any $\beta \in \mathcal{B}$ let $\mathbf{A}(\beta)$ be a $p\times(p-1)-$matrix with real entries such that the $p\times p-$matrix $(\beta\ \mathbf{A}(\beta))$ is orthogonal. The orthogonality is not necessary, invertibility suffices, but orthogonality is expected to lead to better finite sample properties for the test. Let $G(\cdot)$ be an univariate kernel with positive Fourier Transform on the real line. For some bandwidth $\mathfrak{b}$ and any $1 \leq i \neq j \leq n$, let

$$G_{ij}(\beta,\mathfrak{b}) = G((X_i - X_j)^\top \beta/\mathfrak{b}) \quad \text{and} \quad \phi_{ij}(\beta) = \exp(-\|(X_i - X_j)^\top \mathbf{A}(\beta)\|^2/2).$$

Next, with $\widehat{U}_i$ and $\widehat{U}_j$ defined as in equation (IV.4), consider

$$Q_n(\beta) \;=\; \frac{1}{n\,(n-1)\,\mathfrak{b}} \sum_{1\leq i\neq j\leq n} \left\langle \widehat{U}_i(\cdot,\cdot,\beta),\ \widehat{U}_j(\cdot,\cdot,\beta)\right\rangle_n G_{ij}(\beta,\mathfrak{b})\phi_{ij}(\beta),$$

where for any $u(\cdot,\cdot)$ and $v(\cdot,\cdot)$ bounded functions defined on $\mathbb{R}\times\{0,1\}$,

$$\langle u(\cdot,\cdot),\ v(\cdot,\cdot)\rangle_n = \int_{\mathbb{R}\times\{0,1\}} u(t,d)v(t,d)d\widehat{F}_{n,Y,d}(t,d) = \frac{1}{n}\sum_{l=1}^{n} u(Y_l,\delta_l)v(Y_l,\delta_l).$$

The rationale for considering $Q_n(\beta)$ is that it represents a sample based approximation of

$$Q(\beta,\mathfrak{b}) = \mathbb{E}\left[\left\{\int_{\mathbb{R}\times\{0,1\}} U_1(t,d;\beta)U_2(t,d;\beta)dF_{Y,d}(t,d)\right\}\mathfrak{b}^{-1}G((X_1-X_2)^\top\beta/\mathfrak{b})\phi_{12}(\beta)\right].$$

The Inverse Fourier Transform argument used to justify the equation (IV.3) implies that for any $\mathfrak{b} > 0$, $Q(\beta, \mathfrak{b}) \geq 0$ and $Q(\beta, \mathfrak{b}) = 0$ if and only if

$$\mathbb{E}[U(t, d; \beta) \mid X] = 0, \;\; a.s., \;\; \forall (t, d) \in \mathbb{R} \times \{0, 1\}.$$

Given the equivalence from the equation (IV.1) above, under the null hypothesis (III.1) the quantity $Q_n(\beta)$ should be close to zero if β is close to β_0. If the single-index assumption (III.1) does not hold, $Q_n(\beta)$ is expected to be uniformly bounded away from zero. Let us point out that one could obtain $Q_n(\beta)$ from $\widehat{I}_n(\beta)$ after replacing ω_{ij} by $[n/(n-1)]\mathfrak{b}^{-1} G_{ij}(\beta, \mathfrak{b})\phi_{ij}(\beta)$. The univariate smoothing induced by this modification will allow to end with a pivotal test statistics.

The variance of $Q_n(\beta)$ could be estimated by

$$\widehat{\omega}_n(\beta)^2 = \frac{2}{n^2\,(n-1)^2\,\mathfrak{b}^2} \sum_{1 \leq i \neq j \leq n} \left\langle \widehat{U}_i(\cdot, \cdot, \beta), \; \widehat{U}_j(\cdot, \cdot, \beta) \right\rangle_n^2 G_{ij}^2(\beta, \mathfrak{b}) \phi_{ij}^2(\beta).$$

If $\widehat{\beta}$ is the estimator defined in equation (IV.5), then the test statistic we propose is

$$T_n = \frac{Q_n(\widehat{\beta})}{\widehat{\omega}_n(\widehat{\beta})}. \tag{IV.6}$$

By minor modifications of a result in Maistre & Patilea (2014), under some regularity conditions and a suitable rate of decrease to zero for $\mathfrak{b}$, one could show that T_n converges in law to a standard normal distribution if the single-index assumption (III.1) holds true. Moreover, the test could detect directional alternatives of the form

$$\mathbb{P}(Y \leq t, \delta = d \mid X) = \mathbb{P}(Y \leq t, \delta = d \mid X^\top \beta_0) + r_n \delta(X, t, d), \;\; (t, d) \in \mathbb{R} \times \{0, 1\},$$

as soon as $r_n^2 n \mathfrak{b}^{1/2} \to \infty$.

IV.2.2 Confidence regions

One could be interested in confidence regions for the index β_0 and the values of the conditional distribution $F_{T,\beta_0}(\cdot \mid z)$. This could be derived conveniently using numerical methods.

Following Lavergne & Patilea (2013) and Li & Patilea (2014), one could build a randomly perturbed version of $I(\beta)$ as

$$\widehat{I}_n^*(\beta) = \frac{1}{n^2} \sum_{1 \leq i,j \leq n} \left[\frac{1}{n} \sum_{1 \leq l \leq n} \widehat{U}_i(Y_l, \delta_l; \beta) \widehat{U}_j(Y_l, \delta_l; \beta) \right] \omega_{ij}^*, \tag{IV.7}$$

where

$$\omega_{ij}^* = \omega(X_i - X_j)\zeta_i \zeta_j,$$

with $\zeta_1, \cdots, \zeta_n$ a sample of independent exponential random variables with mean equal to 1, independent of the observations. A new estimate of β_0 could be obtained as

$$\widehat{\beta}^* = \arg\min_{\beta\in\mathcal{B}} \widehat{I}_n^*(\beta).$$

The procedure could repeated many times and confidence regions could be derived using the sample of $\widehat{\beta}^*$'s. See, for instance, Chiang & Huang (2012), Lavergne & Patilea (2013), Li & Patilea (2014) for some theoretical justification of this type of approach for building confidence regions. A method for building confidence intervals for $F_{T,\beta_0}(\cdot \mid x^\top\beta_0)$ will be mentioned in the next section.

V Semiparametric estimators of the Kaplan-Meier functionals

With at hand an estimate $\widehat{\beta}$ of the vector β_0 and nonparametric estimates of $H_{0,\beta}(\cdot \mid x^\top\beta)$ and $H_{1,\beta}(\cdot \mid x^\top\beta)$ for arbitrary β, one could use the sample-based version of the formula (III.6) to build an estimator for the Kaplan-Meier functional $F_{T,\beta_0}(\cdot \mid x^\top\beta_0)$. The result will be precisely the conditional Kaplan-Meier estimator with univariate covariate $X^\top\widehat{\beta}$.

Following Beran (1981) and Dabrowska (1989), here we will use Nadaraya-Watson estimators for the sub-distributions of the observations. That is, for any β and $z \in \mathbb{R}$, let

$$\widehat{F}_{T,\beta}((t,\infty] \mid z) = \prod_{Y_i \leq t} \left(1 - \frac{w_{in}(z;\beta)}{\sum_{j=1}^n \mathbf{1}_{\{Y_j \geq Y_i\}} w_{jn}(z;\beta)}\right)^{\delta_i}, \qquad \text{for } t \leq Y_{(n)},$$

where $Y_{(n)} = \max_{1\leq i\leq n} Y_i$ and

$$w_{in}(z;\beta) = \frac{K((X_i^\top\beta - z)h^{-1})}{\sum_{l=1}^n K((X_l^\top\beta - z)h^{-1})},$$

$K(\cdot)$ is a kernel and h is the bandwidth. If $Y_{(n)}$ is an uncensored observation then $\widehat{F}_{T,\beta}((Y_{(n)},\infty] \mid z) = 0$, but when $Y_{(n)}$ is a censored observation, then

$$\widehat{F}_{T,\beta}([Y_{(n)},\infty] \mid z) = \widehat{F}_{T,\beta}((Y_{(n)},\infty] \mid z) > 0.$$

Under the single-index assumption (III.4), the conditional distribution function $F_T(\cdot \mid x)$ of the lifetime T given $X = x$ is equal to $F_{T,\beta_0}(\cdot \mid x^\top\beta_0)$. Hence, a natural estimator of $F_T(\cdot \mid x)$ is $F_{T,\widehat{\beta}}(\cdot \mid x^\top\widehat{\beta})$. Below, we derive a i.i.d. representation for this estimator.

V.1 Asymptotic results for Kaplan-Meier functionals

Let

$$\tau_H(x;\beta_0) = \inf\{t : H((t,\infty) \mid x^\top\beta_0) = 0\},$$

and consider $\overline{\mathcal{X}} \subset \mathcal{X}$ a compact subset in the space of the covariates such that

$$\inf_{x\in\overline{\mathcal{X}}} f_{\beta_0}(x^\top\beta_0) > 0.$$

Let $\tau > 0$ such that

$$\tau < \inf_{x\in\overline{\mathcal{X}}} \tau_H(x;\beta_0).$$

The following i.i.d. asymptotic representation is en extension of the results of Du & Akritas (2002) and Lopez (2011).

Proposition 5.1 *Assume that the single-index condition (III.1) holds true. For each $x \in \overline{\mathcal{X}}$ and $t \leq \tau$, let*

$$\eta_{\Lambda_T,i}(t,x^\top\beta_0) = w_{in}(x^\top\beta_0,\beta_0)\left(\frac{\delta_i\mathbf{1}\{Y_i\leq t\}}{H_{\beta_0}([Y_i,\infty)\mid x^\top\beta_0)} - \int_{(-\infty,t]}\frac{\mathbf{1}\{Y_i\geq s\}H_{1,\beta_0}(ds\mid x^\top\beta_0)}{H^2_{\beta_0}([Y_i,\infty)\mid x^\top\beta_0)}\right)$$

and

$$\begin{aligned}\eta_{F_T,i}(t,x^\top\beta_0) \;=\; -F_{T,\beta_0}((t,\infty]\,|\,x^\top\beta_0)w_{in}(x^\top\beta_0,\beta_0)\Bigg[&\frac{\delta_iF_{T,\beta_0}([Y_i,\infty]\,|\,x^\top\beta_0)\mathbf{1}\{Y_i\leq t\}}{F_{T,\beta_0}((Y_i,\infty]\,|\,x^\top\beta_0)H_{\beta_0}([Y_i,\infty)\,|\,x^\top\beta_0)}\\ &-\int_{(-\infty,t]}\frac{F_{T,\beta_0}([Y_i,\infty]\mid x^\top\beta_0)\mathbf{1}\{Y_i\geq s\}H_{1,\beta_0}(ds\mid x^\top\beta_0)}{F_{T,\beta_0}((Y_i,\infty]\mid x^\top\beta_0)H^2_{\beta_0}([Y_i,\infty)\mid x^\top\beta_0)}\Bigg]\end{aligned}$$

Let the kernel $K(\cdot)$ be a symmetric probability density function with compact support and twice continuously differentiable. Assume that for any $b_n \to 0$ and $d = 0$ or $d = 1$,

$$\sup_{0<\|\beta-\beta_0\|\leq b_n}\left\{\frac{|f_\beta(z)-f_{\beta_0}(z)|}{\|\beta-\beta_0\|} + \frac{|H_{d,\beta}((-\infty,t]\mid z)f_\beta(z) - H_{d,\beta_0}((-\infty,t]\mid z)f_{\beta_0}(z)|}{\|\beta-\beta_0\|}\right\}\leq C$$

where C is a constant independent of $z,t\in\mathbb{R}$. Let $\widehat{\beta}$ be a consistent estimator of β_0. Under the Assumption VIII.1,

$$\begin{aligned}\widehat{\Lambda}_{T,\widehat{\beta}}((-\infty,t]\mid x^\top\widehat{\beta}) - \Lambda_{T,\beta_0}((-\infty,t]\mid x^\top\beta_0) &= \frac{1}{n}\sum_{i=1}^n \eta_{\Lambda_T,i}(t,x^\top\beta_0) + R_{n,\Lambda_T}(t,x)\\ \widehat{F}_{T,\widehat{\beta}}((t,\infty]\mid x^\top\widehat{\beta}) - F_{T,\beta_0}((t,\infty]\mid x^\top\beta_0) &= \frac{1}{n}\sum_{i=1}^n \eta_{F_T,i}(t,x^\top\beta_0) + R_{n,F_T}(t,x),\quad \text{(V.1)}\end{aligned}$$

with

$$\sup_{t\leq\tau,\,x\in\overline{\mathcal{X}}}\{|R_{n,\Lambda_T}(t,x)| + |R_{n,F_T}(t,x)|\} = O_{\mathbb{P}}(n^{-1}h^{-1}\ln n + \|\widehat{\beta}-\beta_0\|).$$

If $\widehat{\beta}$ is a $\sqrt{n}-$estimator, the reminders R_{n,Λ_T}, R_{n,F_T} have the uniform rate $O_{\mathbb{P}}(n^{-1/2})$, that is negligible compared to $O_{\mathbb{P}}(n^{-1/2}h^{-1/2})$, the rate of the i.i.d. sums.

The i.i.d. asymptotic representation is a powerful result that serves to develop asymptotic theory in various situations. See, for instance, Du & Akritas (2002) and Lopez (2011) for some examples and references. Here, we illustrate a direct consequence for a single-index cure rate estimator that extends the result of Xu & Peng (2014). Cure models received a lot of attention lately, see, for instance, Tsodikov *et al.* (2003) and Zheng *et al.* (2006). The conditional cure rate $\pi(x) = \mathbb{P}(T = \infty \mid x)$ could be estimated by the conditional Kaplan-Meier estimator of Beran (1981) taken at the largest uncensored observation. In the case of a univariate fixed-design covariate, Xu & Peng (2014) assume that

$$-\infty < \sup_x \tau_{H_1}(x) < \inf_x \tau_H(x) < \infty,$$

where

$$\tau_{H_1}(x) = \inf\{t \mid H_1((t,\infty) \mid x) = 0\}, \quad \tau_H(x) = \inf\{t \mid H((t,\infty) \mid x) = 0\},$$

so that for any x,

$$\int_{[0,\tau_{H_1}(x)]} \frac{H_1(dt \mid x)}{H^2([t,\tau_H(x)] \mid x)} < \infty.$$

Under these conditions and some technical assumptions, Xu and Peng proved that, for any $t \in [0, \tau_H(x)]$,

$$\sqrt{nh}\left(\widehat{\Lambda}_T([0,t] \mid x) - \Lambda_T([0,t] \mid x)\right) \rightsquigarrow N(0, \sigma^2(t,x)),$$

where

$$\sigma^2(t,x) = \int_{[0,t]} \frac{\Lambda_T(ds \mid x)}{H([s,\tau_H(x)] \mid x)} \int K^2(t)dt$$

and $\rightsquigarrow$ denotes convergence in law. Moreover,

$$\sqrt{nh}\left(\widehat{F}_T((Y^1_{(n)},\infty]) \mid x) - \pi(x)\right) \rightsquigarrow N(0, \pi^2(x)\sigma^2(x)),$$

where $\widehat{F}_T$ is the Beran estimator, $Y^1_{(n)}$ is the largest uncensored observation of Y and

$$\sigma^2(x) = \int_{[0,\tau_{H_1}(x)]} \frac{\Lambda_T(ds \mid x)}{H([s,\tau_H(x)] \mid x)} \int K^2(t)dt.$$

Our i.i.d. representation allows to recover this result and to extend it to multivariate covariates using the single-index framework, as stated in the following corollary.

Corollary 5.2 *Assume that the assumptions of Proposition 5.1 and the assumptions (III.3) and (III.4) hold true. Then*

$$\pi(x) = \mathbb{P}(T = \infty \mid x) = F_{T,\beta_0}((\tau_H(x), \infty] \mid x^\top \beta_0), \qquad \forall x \in \mathcal{X}.$$

Assume

$$-\infty < \sup_{x \in \mathcal{X}} \tau_{H_1}(x; \beta_0) < \inf_{x \in \overline{\mathcal{X}}} \tau_H(x; \beta_0) < \infty,$$

where $\tau_{H_1}(x; \beta_0) = \inf\{t \mid H_{1,\beta_0}((t, \infty) \mid x^\top \beta_0) = 0\}$. *If* $x \in \overline{\mathcal{X}}$ *and* $\widehat{\beta}$ *is a* $\sqrt{n}-$*consistent estimator of* β_0*,*

$$\sqrt{nh}\left(\widehat{F}_{T,\widehat{\beta}}((Y^1_{(n)}, \infty]) \mid x^\top \widehat{\beta}) - \pi(x)\right) \rightsquigarrow N(0, \pi^2(x)\sigma^2_{\beta_0}(x^\top \beta_0))$$

$$\sigma^2_{\beta_0}(x) = \int_{(-\infty, \tau_{H_1}(x;\beta_0)]} \frac{H_{1,\beta_0}(ds \mid x^\top \beta_0)}{H^2_{\beta_0}([s, \tau_H(x)] \mid x^\top \beta_0)} \int K^2(t) dt.$$

To build confidence regions for $F_{T,\beta_0}((-\infty, t] \mid x^\top \beta_0)$ for given x and t, one could resample (Y, δ) from the subdistributions $H_{1,\widehat{\beta}}$ and $H_{0,\widehat{\beta}}$. This could be done under a fixed design assumption, as Van Keilegom & Veraverbeke (1997) did in the case of a univariate covariate, or taking into account a univariate random design as in Li & Datta (2001). We argue that given the parametric convergence rate of $\widehat{\beta}$, the validity of such bootstrap procedures could be derived by similar arguments as used in the univariate case. The detailed investigation of this issue will be considered elsewhere.

VI Empirical evidence

To illustrate the new approach, we performed and present in the following few empirical experiments with simulated and real data.

VI.1 Simulation experiments

We consider a lifetime T defined as $T = 5(X^\top \beta_0)^2 + \varepsilon$ with a three-dimensional vector of covariates $X = (X^{(1)}, X^{(2)}, X^{(3)})^\top$, where $(X^{(1)}, X^{(2)})^\top$ are generated from a zero mean bivariate normal with standard deviation equal to 1 and correlation equal to 0.2, and $X^{(3)}$ is a Bernoulli random variable with parameter $p = 0.4$. The variable ε is standard normal and independent of X. The true value of the parameter is $\beta_0 = (\beta_{01}, \beta_{02}, \beta_{03})^\top = (1, 1, 1)^\top$. The censoring variable is $C = c\ \eta\ \exp(\sqrt{|X^\top \beta_0|/5}) - 3$ with η uniformly distributed on $[0, 1]$ and c a constant which controls the proportion of censoring. In our simulation, we use $c = 5.8$ and 20.3 which yields approximately 80% and 40% censoring proportion, respectively. Let us recall that in order to guarantee identification, we set the first component of $\beta = (\beta_1, \beta_2, \beta_3)^\top$ equal to 1.

First, we proceed to the estimation of the index parameter β_0 using the estimator $\widehat{\beta} = (1, \widehat{\beta}_2, \widehat{\beta}_3)^\top$ proposed in equation (IV.5). We compare our method to the hMAVE (MAVE for hazard function) proposed by Xia *et al.* (2010). They prefer to identify the parameter β by constraining the norm to be equal to 1 (and fixing the sign of one component). Here, in order to have comparable results, we divide the estimate of β obtained with their method by its first component. (We also performed simulations, not reported here, using Xia *et al.* (2010)'s identification condition in our equation (IV.5). The conclusions are very similar.) To compare performances, we report several statistics obtained with 500 independent samples of size $n = 50$: the mean, the standard deviation and the median of the estimators of β_2 and β_3, as well as the mean and the standard deviation of the largest singular value (sv) of $\widehat{\beta}\widehat{\beta}^\top - \beta_0\beta_0^\top$ and of the absolute estimation errors (aee) $|\widehat{\beta}_2 - \beta_{02}| + |\widehat{\beta}_3 - \beta_{03}|$. The kernel $L(\cdot)$, which is used to define $\widehat{U}_i(t, d; \beta)$ in equation (IV.4), is the gaussian kernel. The bandwidth for this kernel is selected as the minimizer of the loss $\widehat{I}_n(\widehat{\beta})$ with respect to g on an equidistant grid $\{0.01, 0.03, \cdots, 0.13\}$. The simulation results are shown in Table 1. Our method performs quite well. Compared with hMAVE, it has a similar absolute bias, but the law of our estimator seems more concentrated.

Table 1: Descriptive statistics for the estimators of $\beta_0 = (1, 1, 1)^\top$, and the mean and the standard deviation (in parentheses) of the *aee* and *sv* values, obtained with 500 independent samples of size $n = 50$. The results obtained with the method of Xia *et al.* (2010) are presented in the gray cells.

	80% censoring					
	mean		std		median	
β_2	1.0539	0.9838	0.2902	0.4214	0.9876	0.9920
β_3	0.9819	0.9435	0.2383	0.5029	0.9724	0.9678
aee	0.2985(0.4000)			0.6260(0.5430)		
sv	0.6730(1.1801)			1.3661(1.3603)		
	40% censoring					
	mean		std		median	
β_2	0.9955	1.0011	0.1934	0.2249	0.9691	0.9816
β_3	0.9813	0.9844	0.1608	0.2890	0.9735	0.9737
aee	0.2133(0.2529)			0.3960(0.2547)		
sv	0.4584(0.7236)			0.8828(0.6637)		

Next, we used the idea described in section IV.2.2 above to build confidence intervals for β_{02} and β_{03}. For each of the 500 samples of size $n = 50$ we generated 299 independent random samples $\zeta_1, \cdots, \zeta_n$ and computed the criteria $\widehat{I}_n^*(\beta)$ as in equation (IV.7). Only the setup with 80% of censoring is reported. The 90% and 95% confidence intervals obtained with the optimal values $\widehat{\beta}_2^*$ and $\widehat{\beta}_3^*$ are presented in Table 2. The levels for β_{02} are slightly less than the nominal ones, while the intervals for the coefficient of the discrete component of X are slightly larger than necessary. Overall, the approach we propose for building confidence intervals performs reasonably well with small samples where the asymptotic approximation could be quite poor.

Table 2: Small sample confidence intervals: a number of 299 perturbed criteria $\widehat{I}_n^*(\beta)$ as defined in equation (IV.7) are used for any of the 500 samples of size $n = 50$. The censoring amount is 80%.

	90% confidence interval		95% confidence interval	
	length	percentage	length	percentage
β_2	0.2110	88.4	0.2758	94.2
β_3	0.1955	93.4	0.2588	97.6

Finally, we considered the estimation of $F_{T,\beta_0}((t,\infty] \mid x) = F_{T,\beta_0}((t,\infty] \mid x^\top\beta_0)$ using the conditional Kaplan-Meier estimator. The values of t considered are the quantiles 0.1, 0.2,...,0.9 of the true conditional law. To approximate β_0 we considered both cases, our estimator and Xia *et al.* (2010)'s estimator. We fixed two vectors x, namely $x = (-0.5,-0.5,1)^\top$ and $x = (0.1,-0.3,0)^\top$, for which we considered the respective amount of censoring of 80% and 40%. The quartic (biweight) kernel $K(z) = (15/16)(1-z^2)^2\mathbf{1}\{|z| \leq 1\}$ is used, and the bandwidth h is taken equal to 0.4 with each of the two estimates of β_0 we use. (The quartic kernel is not twice differentiable as required in our Proposition 5.1, but since our empirical experience shows that this incoherence has practically no impact, we keep it for reasons of comparison with previous contributions in the literature.) The results are presented in Tables 3 and 4, those obtained with Xia *et al.*'s estimate of β_0 are again in gray cells. The estimates based on our $\widehat{\beta}$ are less biased and overall quite accurate.

Table 3: Single-index estimator of the conditional distribution: 500 samples of size $n = 50$, censored rate is 80%, $x = (-0.5,-0.5,1)^\top$

$F_{T,\beta_0}((t,\infty] \mid x^\top\beta_0)$	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
$\widehat{F}_{T,\widehat{\beta}}((t,\infty] \mid x^\top\widehat{\beta})$	0.921	0.838	0.757	0.670	0.587	0.497	0.400	0.301	0.180
	0.943	0.883	0.819	0.742	0.669	0.594	0.506	0.410	0.295

Table 4: Single-index estimator of the conditional distribution: 500 samples of size $n = 50$, censored rate is 40%, $x = (0.1,-0.3,0)^\top$

$F_{T,\beta_0}((t,\infty]\|x^\top\beta_0)$	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
$\widehat{F}_{T,\widehat{\beta}}((t,\infty]\|x^\top\widehat{\beta})$	0.906	0.819	0.729	0.631	0.544	0.455	0.361	0.261	0.153
	0.906	0.830	0.745	0.650	0.573	0.491	0.398	0.298	0.187

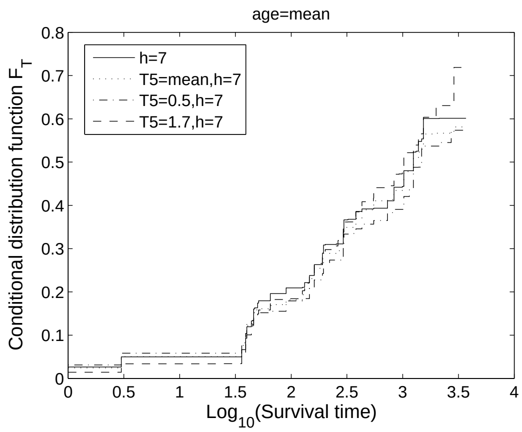
VI.2 Real data application

For a real data illustration we consider the well-known Stanford heart transplantation data set given by Miller & Halpern (1982). The data set includes survival times and covariates for 184 patients who had received a heart transplantation between October 1967 and February 1980. For reasons of comparison with previous empirical work, see for instance Miller and Halpern (1982), Wei *et al.* (1990) and van Keilegom *et al.* 2001), we restrict our attention to the 157 out of 184 individuals who had complete tissue typing. Moreover, the response variable will be taken equal to the base 10 logarithm of the survival time (in days). The covariates will be the age of the patient at the time of the first heart transplant and the T5 mismatch score which measure the degree of tissue incomputability between the initial donor and recipient hearts with respect to HLA antigens. A set of 55 patients alive beyond February 1980 were considered censored.

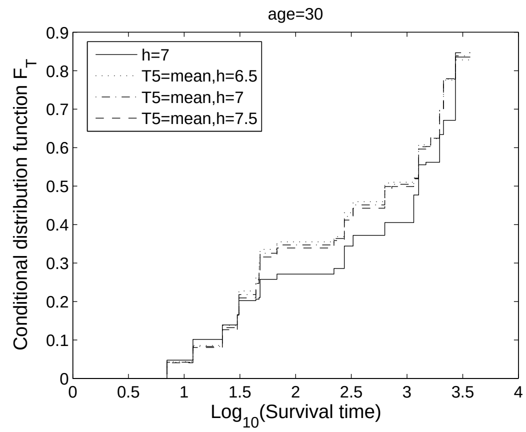
For identification with our single-index approach, the coefficient of the covariate 'age' is fixed equal to 1. The estimate of the coefficient of the covariate 'T5 mismatch score' is equal to 2.9333. This value may seem surprising given the previous analysis in the literature, but one has to keep in mind the normalization induced by the identification assumption. The 95% confidence interval is (0.2972, 3.0769) which suggests that this covariate could be significant. The effect of the T5 mismatch score is not captured by the classical models which indicate that the coefficient of this covariate is not significant (see for instance Table 2 in Miller & Halpern (1982) and Table 1 in Wei *et al.* (1990)). Next, we applied the conditional Kaplan-Meier estimator with the index $(1, 2.9333)^\top$ and several values of the covariates, using the quartic kernel. The estimated curves are presented in the four panels of Figure 1. In panel (a) we considered the bandwidth $h = 7$, as considered by van Keilegom *et al.* (2001), and we set the covariate 'age' equal to its empirical mean. The curve corresponding to a null value for the covariate 'T5 mismatch score', as well as the curves corresponding to three non-null values of the covariate 'T5 mismatch score' are presented. The differences between the four curves advocate for an effect of the T5 mismatch score. In panel (b) we investigate the effect of a change in the bandwidth h when 'age' is equal to 30 and 'T5 mismatch score' is equal to its empirical mean. Quite little change could be noticed when h varies from 6.5 to 7 and to 7.5. Once again, the effect of 'T5 mismatch score' seems significant in view of the curve corresponding to the absence of this covariate, also plotted in the panel (b). The experience from panel (b) was repeated with 'age' equal to 40 and 50 and the results are presented in the panels (c) and (d), respectively. The same conclusion could be drawn: the effect of a slight change in the bandwidth h seems small. Meanwhile, the effect of the covariate 'T5 mismatch score' could be more important, suggesting that this covariate is significant.

VII Discussion and extensions

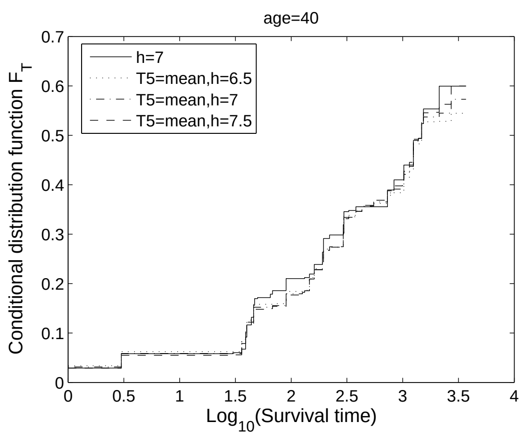
We propose a new way to build estimators under dimension reduction assumptions in survival analysis with covariates. The main idea is to impose the dimension reduction as-



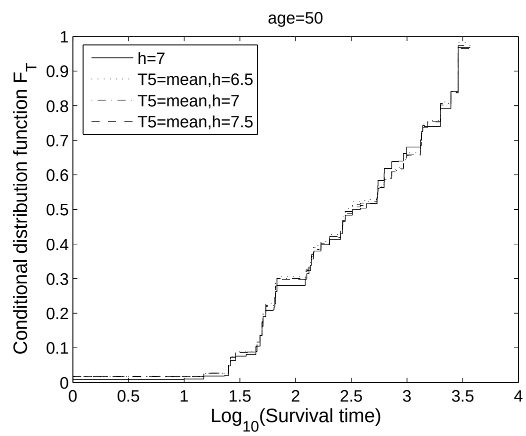
(a)



(b)



(c)



(d)

Figure 1: The estimates of the conditional distribution function. The continuous lines correspond to the estimates obtained using with only the covariate 'age', as considered by van Keilegom *et al.* (2001).

sumption in the observations space and to estimate the conditional law of the observations under such conditions. Here, we consider a well-known dimension reduction approach based on indices, that is we assume that all the information on the responses carried by the covariates is contained in a small number of linear combinations of them. These linear combinations of the covariates is given by unknown index vectors. It is worthwhile to note that such a dimension reduction introduced through the observations space could be more easily tested, for instance following the approach introduced by Maistre & Patilea (2014). Next, the idea is to use the map generated by the censoring mechanism that links the conditional law of the observations to the quantities of interest and to plug-in the estimates of the conditional law of the observations. As a result, we obtain easy to calculate semiparametric estimates of the conditional law of the lifetime of interest and of any other smooth functional of the observations. Moreover, bootstrapping in the space of the observations and applying the map to the conditional law estimate obtained with each bootstrap sample, we could easily build confidence regions for the quantities of interest. In this paper, we focus on single-index assumptions and random right censoring but our principle is much more general, as we explain below. The asymptotic properties of the estimators could be rapidly derived and under mild conditions. We also showed that the related contributions in the literature eventually lead to a single-index structure on the conditional law of the observations, like the one we consider herein. However, the existing estimation procedures are much more complicated due to the fact that they involve simultaneously the two aspects: dimension reduction and the map generated by the censoring mechanism linking the observations to the quantities of interest.

Let us now suggest some possible extensions. Other asymptotic results already proved for the conditional Kaplan-Meier approach could be adapted to the dimension reduction idea to derive single-index versions of them. For instance, one could consider a single-index version of the Bahadur-type representation derived by Dabrowska (1992) for the estimator of the conditional quantile of T obtained by inversion of the classical conditional Kaplan-Meier estimator.

Clearly, our dimension reduction approach could be extended to multi-index assumptions. Suppose there exists a $p \times r-$matrix B_0 with $1 \leq r < p$ such that

$$(Y, \delta) \perp X \mid X^\top B_0. \tag{VII.1}$$

A similar condition is used by Xia *et al.* (2010). To estimate a basis in the space generated by the columns of B_0, it suffices to redefine the $\widehat{U}_i$ introduced in equation (IV.4) as

$$\widehat{U}_i(t, k; B) = \frac{1}{n-1} \sum_{k=1}^{n} \{\mathbf{1}\{Y_i \leq t; \delta = d\} - \mathbf{1}\{Y_k \leq t; \delta = d\}\} \frac{1}{g^r} L_{ik}(B, g),$$

with $L_{ik}(B, g) = L((X_i - X_k)^\top B/g)$ and $L(\cdot)$ a $r-$dimension kernel. Next, define $I_n(B)$ accordingly and minimize $I_n(B)$ with respect cu B under suitable identification restriction (for instance, the top block of B could be fixed equal to the $r \times r$ identity matrix). As pointed out by Maistre & Patilea (2014), the test statistic T_n defined in equation (IV.6) could be modified in order to test the null hypothesis (VII.1) of a $r-$dimension multi-index hypothesis. With at hand an estimate $\widehat{B}$, one can build the conditional Kaplan-Meier estimator with $r-$dimension covariate vectors $X_i^\top \widehat{B}$, just like in Beran (1981) and Dabrowska (1989).

Finally, let us point out that in principle our approach could apply to a larger class of censoring mechanisms. The implementation and the theoretical investigation of the estimators is much easier when the map linking the conditional law of the observations to the quantities of interest has a closed form expression. Let us end with another example of censoring mechanism considered by Patilea & Rolin (2006) that generates a closed form expression map between the observations and a lifetime of interest $T \in [0, \infty]$. The observations are independent copies of (Y, A, X) where X is a vector of p covariates, $A \in \{0, 1, 2\}$, $Y \in [0, \infty)$ and

$$\begin{cases} Y < T & \text{if} \quad A = 0, \\ Y = T & \text{if} \quad A = 1, \\ Y > T & \text{if} \quad A = 2. \end{cases}$$

Such observations could occur for instance when the age of the occurrence of a disease is under study and only one of the following situation is possible: a) evidence of the disease is present and the age at onset is known (from medical records, interviews with the patient or family members,...); b) the disease is diagnosed but the age at onset is unknown or the accuracy of the information about this is questionable; and c) the disease is not diagnosed at the examination time. Following our approach, assume that $(Y, A) \perp X \mid X^\top \beta_0$ for some $\beta_0 \in \mathbb{R}^p$. For $\beta \in \mathbb{R}^p$, $0 \le t < \infty$, define the sub-distributions

$$H_{k,\beta}([0, t] \mid z) = P(Y \le t, A = k \mid X^\top \beta = z), \quad k = 0, 1, 2, \quad z \in \mathbb{R}.$$

The conditional law of the observations is given by $H_{k,\beta_0}(\cdot \mid z)$, $k = 0, 1, 2$. Let C be a censoring time and $\Delta \in \{0, 1\}$ a 'latent' variables that models the fact that T is observed or not when $T \le C$. That means

$$Y = \min(T, C) + (1 - \Delta) \max(C - T, 0) = C + \Delta \min(T - C, 0)$$

and $A = 2(1 - \Delta)\mathbf{1}\{T \le C\} + \mathbf{1}\{C < T\}$. Let $p_\beta(z) = \mathbb{P}(\Delta = 1 \mid X^\top \beta = z)$, $z \in \mathbb{R}$. If T, C and Δ are independent given X, for any z we can write

$$\begin{cases} H_{0,\beta_0}(dt \mid z) &= F_{T,\beta_0}((t, \infty] \mid z) F_{C,\beta_0}(dt \mid z), \\ H_{1,\beta_0}(dt \mid z) &= p_{\beta_0}(z) F_{C,\beta_0}([t, \infty] \mid z) F_{T,\beta_0}(dt \mid z), \\ H_{2,\beta_0}(dt \mid z) &= (1 - p_{\beta_0}(z)) F_{T,\beta_0}([0, t] \mid z) F_{C,\beta_0}(dt \mid z). \end{cases}$$

The case $p \equiv 1$ corresponds to the usual right-censoring setup, while $p \equiv 0$ corresponds to current-status data. If $p_{\beta_0}(z) > 0$ the system could be solved explicitly to obtain

$$p_{\beta_0}(z) = \frac{H_{0,\beta_0}([0, \infty) \mid z)}{H_{0,\beta_0}([0, \infty) \mid z) + H_{2,\beta_0}([0, \infty) \mid z)}$$

and

$$\Lambda_{T,\beta_0}(dt \mid z) = \frac{H_{0,\beta_0}(dt \mid z)}{H_{0,\beta_0}([t,\infty) \mid z) + p_{\beta_0}(z) H_{1,\beta_0}([t,\infty) \mid z)}.$$

Moreover, one could allow for a positive probability of the event $\{T = \infty\}$ and write

$$\mathbb{P}(T = \infty \mid X^\top \beta_0 = z) = \prod_{t \in (0,\infty)} \{1 - \Lambda_{T,\beta_0}(dt \mid z)\}.$$

An estimate $\widehat{\beta}$ for β_0 could be build by an obvious extension of the method proposed in section IV.1. Then one can easily estimate $H_{k,\widehat{\beta}}([0,t] \mid x^\top \widehat{\beta})$, $k = 0, 1, 2$, by kernel smoothing and plug-in the estimates in the formulae above to obtain estimates for the quantities of interest.

References

[1] Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical report, Univ. California, Berkeley.

[2] Bouaziz, O. & Lopez, O. (2010). Conditional density estimation in a censored single-index model. *Bernoulli* **16**, 514–542.

[3] Chiang, C.-T. & Huang, M.-Y. (2012). New estimation and inference procedures for a single-index conditional distribution model. *J. Multivariate Anal.* **111**, 271–285.

[4] Cox, D.R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187–220.

[5] Dabrowska, D. M. (1989). Uniform Consistency of the Kernel Conditional Kaplan-Meier Estimate. *Ann. Statist.* **17**, 1157–1167.

[6] Dabrowska, D. M. (1992). Nonparametric quantile regression with censored data. *Sankhyā* **54**, 252–259.

[7] Delecroix, M., Härdle, W. & Hristache, M. (2003). Efficient estimation in conditional single-index regression. *J. Multivariate Anal.* **86**, 213–226.

[8] Du, Y. & Akritas, M. G. (2002). I.i.d representations of the conditional KaplanMeier process for arbitrary distributions. *Math. Meth. Statist.* **11**, 152-182.

[9] Einmahl, U. & Mason, D.M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.* **33**, 1380–1403.

[10] Gill, R.D. & Johansen, S. (1990). A Survey of Product-Integration with a View Toward Application in Survival Analysis. *Ann. Statist.* **18**, 1501–1555.

[11] Hall, P. & Yao, Q. (2005). Approximating conditional distribution functions using dimension reduction. *Ann. Statist.* **33**, 1404–1421.

[12] Horowitz, J. (2009). *Semiparametric and Nonparametric Methods in Econometrics.* Springer Series in Statistics. Springer: New-York.

[13] Lavergne, P. & Patilea, V. (2013). Smooth Minimum Distance Estimation and Testing in Conditional Moment Restrictions Models: Uniform in Bandwidth Theory. *J. Econometrics* **177**, 47–59.

[14] Lee, K.-Y., Li, B. & Chiaromonte, F. (2013). A general theory for nonlinear sufficient dimension reduction: Formulation and estimation. *Ann. Statist.* **41**, 221–249.

[15] Li, G. & Datta S. (2001). Bootstrap approach to nonparametric regression for right censored data. *Ann. Inst. Statist. Math.* **53**, 708–729.

[16] Li, W. & Patilea V. (2014). A new semiparametric estimator for single-index regressions. Working Paper, CREST-Ensai.

[17] Lopez, O. (2011). Nonparametric estimation of the multivariate distribution function in a censored regression model with applications. *Comm. Statist. Theory Methods* **40(15)**, 2639–2660.

[18] Lu, X. (2010). Asymptotic distributions of two "syntheticdata" estimators for censored single-index models. *J. Multivariate Anal.* **101**, 999–1015.

[19] Ma, Y. & Zhu, L. (2013). Efficient estimation in sufficient dimension reduction. *Ann. Statist.* **41**, 250–268.

[20] Maistre, S. & Patilea, V. (2014). Nonparametric model checks of single-index assumptions. arXiv:1410.4934 [math.ST].

[21] Miller, R. & Halpern, J. (1982). Regression with censored data. *Biometrika* **69**, 521–531.

[22] Patilea, V., & Rolin, J.M. (2006). Product-limit estimators of the survival function for two modified forms of current-status data. *Bernoulli* **12**, 801-819.

[23] Sherman, R.P. (1994). Maximal inequalities for degenerate $U-$processes with applications to optimization estimators. *Ann. Statist.* **22**, 439–459.

[24] Strzalkowska-Kominiak, E. & Cao, R. (2013). Maximum likelihood estimation for conditional distribution single-index models under censoring. *J. Multivariate Anal.* **114**, 74–98.

[25] Strzalkowska-Kominiak, E. & Cao, R. (2014). Beran-based approach for single-index models under censoring. *Comput. Statist.* **29**, 1243–1261.

[26] Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *J. Multivariate Anal.* **45**, 89–103.

[27] Tsodikov, A.D., J. G Ibrahim, J.G. & Yakovlev, A.Y. (2003). Estimating Cure Rates From Survival Data: An Alternative to Two-Component Mixture Models. *J. Amer. Statist. Assoc.* **98(464)**, 1063–1078.

[28] Van Keilegom, I. & Veraverbeke, N. (1996). Uniform strong convergence for the conditional Kaplan-Meier estimator and its quantiles. *Comm. Statist. Theory Methods* **25**, 2251-2265.

[29] Van Keilegom, I. & Veraverbeke, N. (1997). Weak convergence of the bootstrapped conditional Kaplan-Meier process and its quantile process. *Comm. Statist. Theory Methods* **26**, 853–869.

[30] Van Keilegom, I., Akritas, M.G. & Veraverbeke, N. (2001). Estimation of the conditional distribution in regression with censored data: a comparative study. *Comput. Statist. Data Anal.* **35**, 487–500.

[31] Wei, L.J., Ying Z. & Lin, D.Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845–851.

[32] Xia, Y., Dixin Zhang, D. & Xu, J. (2010). Dimension Reduction and Semiparametric Estimation of Survival Models. *J Amer. Statist. Assoc.* **105(489)**, 278–290.

[33] Xu, J., & Peng, Y. (2014). Nonparametric cure rate estimation with covariates. *Canad.J. Statist.* **42**, 1–17.

[34] Zheng, D., Yin, G. & Ibrahim, J.G. (2006). Semiparametric Transformation Models for Survival Data With a Cure Fraction. *J. Amer. Statist. Assoc.* **101(474)**, 670–684.

VIII Appendix

VIII.1 Proof of equation (IV.3)

Here we justify the statement

$$I(\beta) \geq 0 \text{ and } I(\beta) = 0 \quad \text{if and only if} \quad \mathbb{E}[U(t,d;\beta) \mid X] = 0 \;\; a.s. \;\; \forall t \in \mathbb{R}, \; d \in \{0,1\}.$$

Let $\mathcal{F}[\omega](u) = \int_{\mathbb{R}^p} e^{-2\pi i x^\top u} \omega(x) dx$, $u \in \mathbb{R}^p$, denote the Fourier Transform of $\omega(\cdot)$. If $\mathcal{F}[\omega]$ is integrable, by the Inverse Fourier Transform formula and Fubini Theorem, we can write

$$\begin{aligned} I(\beta) &= \int_{\mathbb{R}\times\{0,1\}} \mathbb{E}\left[\omega(X_1 - X_2) U_1(t,d;\beta) U_2(t,d;\beta)\right] d\mu(t,d) \\ &= \int_{\mathbb{R}\times\{0,1\}} \mathbb{E}\left[U_1(t,d;\beta) U_2(t,d;\beta) \int_{\mathbb{R}^p} e^{2\pi i (X_1 - X_2)^\top u} \mathcal{F}[\omega](u) du\right] d\mu(t,d) \\ &= \int_{\mathbb{R}\times\{0,1\}} \int_{\mathbb{R}^p} \left| \mathbb{E}\left[\mathbb{E}\left[U(t,d;\beta) \mid X\right] e^{2\pi i X^\top u}\right]\right|^2 \mathcal{F}[\omega](u) du d\mu(t,d). \end{aligned}$$

The statement follows from the fact that $\mathcal{F}[\omega]$ is positive and the uniqueness of the Fourier Transform.

VIII.2 Assumptions

Let us introduce some notation. Let $\widetilde{X} \in \mathbb{R}^{p-1}$ be the $(p-1)-$dimension vector of the last components of X. Below, $(\widetilde{X})_r$ (resp. $(\widetilde{X}\widetilde{X}^\top)_{rq}$) denotes the rth components (resp. the $rq-$entry) of the vector $\widetilde{X}$ (resp. matrix $\widetilde{X}\widetilde{X}^\top$). If A is a matrix with real entries, $\|A\| = \sqrt{trace(A^\top A)}$. In the following, where ∂_z (resp. ∂^2_{zz}) denotes the first (resp. second) order derivative with respect to z. Some comments on the following assumptions are provided in the Supplementary Material.

Assumption VIII.1 *1. The observations* (Y_i, δ_i, X_i), $1 \leq i \leq n$, *are independent copies of* $(Y, \delta, X) \in \mathbb{R} \times \{0,1\} \times \mathbb{R}^p$. *Moreover, there exists a positive number* a *such that* $\mathbb{E}[\exp(a\|X\|)] < \infty$.

2. *The law of* (Y, δ, X) *is such that* $\int_{\mathbb{R}} [H([t,\infty) \mid x)]^{-1} H_0(dt \mid x) = \infty$, $\forall x$.

3. *The parameter set is* $\mathcal{B} = \{1\} \times \mathcal{B}'$ *and* $\mathcal{B}' \subset \mathbb{R}^{p-1}$ *is a compact set with non-empty interior. The vector* $\beta_0 \in \mathcal{B}$ *satisfying the condition (III.1) is the unique element* $\mathcal{B}$ *having this property and the sub-vector built with its last* $p-1$ *components is in the interior of* $\mathcal{B}'$. *For any* $\beta \in \mathcal{B}$ *the random variable* $X^\top \beta$ *has a density* f_β.

4. *The value* β_0 *is a well-separated point of minimum for* $I(\beta)$ *defined in equation (IV.2) with* $\omega(x) = \exp(-\|x\|^2/2)$ *and* μ *equal to the distribution* $F_{Y,\delta}$ *of the observations* (Y, δ), *that means, for any* $\varepsilon > 0$, $\inf_{\beta \in \mathcal{B}, \|\beta - \beta_0\| \geq \varepsilon} I(\beta) > I(\beta_0)$.

5. $$\sup_{z\in\mathbb{R}} \mathbb{E}\left[\|\widetilde{X}\|^4 \mid X^\top\beta_0 = z\right] f_{\beta_0}(z) < \infty \tag{VIII.2}$$

6. *For each* $d \in \{0,1\}$ *and* $t \in \mathbb{R}$, $0 \le r, q \le p-1$ *the functions*

$$z \mapsto f_{\beta_0}(z), \quad z \mapsto H_{d,\beta_0}((-\infty, t] \mid X^\top\beta_0 = z),$$

$$z \mapsto \mathbb{E}[(\widetilde{X})_r \mid X^\top\beta_0 = z] \quad \text{and} \quad z \mapsto \mathbb{E}[(\widetilde{X}\widetilde{X}^\top)_{rq} \mid X^\top\beta_0 = z]$$

are four times continuously differentiable and the derivatives up to order four are bounded, respectively uniformly bounded with respect to t *in the case of* H_{d,β_0}. *The fourth order derivative are Lipschitz functions, with a Lipschitz constant independent of* t *in the case of* H_{d,β_0}.

7. *Let* A *be the set of values* $(t,d) \in \mathbb{R} \times \{0,1\}$ *such that*

$$Var\left[\left(\widetilde{X} - \mathbb{E}\left[\widetilde{X} \mid X^\top\beta_0\right]\right) \partial_z \left(H_{d,\beta_0}((-\infty,t] \mid \cdot)\right)(X^\top\beta_0)\right] \tag{VIII.3}$$

is positive definite. Then $F_{Y,\delta}(A) > 0$.

8. *Let* $z \mapsto \lambda_\beta(z;t,d)$ *denote any of the four functions at point 6 above and their derivatives up to the second order, considered for any* $\beta \in \mathcal{B}$. *Then, the family of functions* $\{\lambda_\beta(\cdot;t,d) : \beta \in \mathcal{B}, t \in \mathbb{R}, d = 0,1\}$ *is a VC-class (or Euclidian) for an envelope with finite moment of order 8. Moreover, for any sequence* $b_n \to 0$,

$$\sup_{\|\beta-\beta_0\|\le b_n} \sup_{t\in\mathbb{R}, d\in\{0,1\}} \sup_{z\in\mathbb{R}} |\lambda_\beta(z;t,d) - \lambda_{\beta_0}(z;t,d)| \to 0.$$

9. *The kernel* $L(\cdot)$ *is a symmetric and twice continuously differentiable univariate density with the second order derivative with bounded variation. Moreover, for* $\kappa = 1,2$, $\int_{\mathbb{R}} |L^{(\kappa)}(u)|du < \infty$, *where* $L^{(\kappa)}(\cdot)$ *denotes the* κ*th derivative of* $L(\cdot)$.

10. $ng^4 \to 0$ *and* $ng^{3+a} \to \infty$ *for some* $a \in (0,1)$.

VIII.3 Proof of Proposition 4.1

Let us introduce some notation. For $\beta \in \mathcal{B} \subset \{1\} \times \mathbb{R}^{p-1}$ let $\widetilde{\beta} \in \mathbb{R}^{p-1}$ denote the sub-vector of its last $(p-1)$ components. Let ∇_β the differential operator given by the $(p-1)$ last first order partial derivatives corresponding to the last $(p-1)$ components of β. This means, $\nabla_\beta \widehat{I}_n(\beta)$ represents the $(p-1)-$dimension vector of first order partial derivatives of $I_n(\beta)$ with respect to $\widetilde{\beta}$ and $\nabla^2_{\beta\beta} I_n(\beta)$ denotes the corresponding Hessian $(p-1)\times(p-1)-$matrix. For each i, let $\widetilde{X}_i \in \mathbb{R}^{p-1}$ be the $(p-1)-$dimension vector of the last components of X_i. We will also use the following simplified notation :

$$\sup_i = \sup_{1\le i\le n}, \qquad \sup_{t,d} = \sup_{(t,d)\in\mathbb{R}\times\{0,1\}}.$$

In the following, for a sequence W_n, $n \geq 1$ of random vectors, $W_n = O_{\mathbb{P}}(1)$ (resp. $W_n = o_{\mathbb{P}}(1)$) means $\|W_n\| = O_{\mathbb{P}}(1)$ (resp. $\|W_n\| = o_{\mathbb{P}}(1)$).

Proof of Proposition 4.1. Since by assumption β_0 is a well-separated point of minimum for $I(\beta)$, it suffices to prove that

$$\sup_{\beta \in \mathcal{B}} \left| \widehat{I}_n(\beta) - I(\beta) \right| = o_{\mathbb{P}}(1).$$

This uniform convergence follows by Hoeffding decomposition of $U-$statistic of order 4 obtained from $\widehat{I}_n(\beta)$ after removing negligible diagonal terms, and by results on the rate of uniform convergence of degenerate $U-$statistics, like for instance those of Sherman (1994). See also the Maximal Inequality in Supplementary Material. Li & Patilea (2014) provide complete arguments in the case where $\mathbb{P}(\delta = 1) = 1$, our case herein is very similar and hence we omit the details.

First, we prove that $\widehat{\beta} - \beta_0 = O_{\mathbb{P}}(n^{-1/2})$. For this purpose, it suffices to prove

$$\nabla_\beta \widehat{I}_n(\beta_0) = O_{\mathbb{P}}(n^{-1/2}) \tag{VIII.4}$$

and there exists a positive definite matrix $J(\beta_0)$ such that for any sequence of $\overline{\beta}$ between $\widehat{\beta}$ and β_0,

$$\mathbb{P}\left(1/c \leq \|\nabla^2_{\beta\beta} \widehat{I}_n(\overline{\beta}) - J(\beta_0)\| \leq c\right) \to 1, \tag{VIII.5}$$

for some $c > 0$. See for instance Theorem 1-(ii) of Sherman (1994).

Let us write

$$\widehat{U}_i(t, d; \beta) = \frac{1}{n-1} \sum_{k=1}^{n} \left\{ \mathbf{1}\{Y_i \leq t; \delta_i = d\} - \mathbf{1}\{Y_k \leq t; \delta_k = d\} \right\} \frac{1}{g} L_{ik}(\beta, g),$$

where $L_{ik}(\beta, g) = L((X_i - X_k)^\top \beta / g)$, and recall the notation

$$U_i(t, d; \beta) = \left\{ \mathbf{1}\{Y_i \leq t, \delta_i = d\} - H_{d,\beta}((-\infty, t] \mid X_i^\top \beta) \right\} f_\beta(X_i^\top \beta).$$

If

$$\widehat{I}_n(t, d; \beta) = \frac{1}{n^2} \sum_{1 \leq i,j \leq n} \widehat{U}_i(t, d; \beta) \widehat{U}_j(t, d; \beta) \omega_{ij}, \qquad (t, d) \in \mathbb{R} \times \{0, 1\},$$

with $\omega_{ij} = \omega(X_i - X_j)$, then

$$\begin{aligned} \frac{1}{2} \nabla_\beta \widehat{I}_n(t, d; \beta_0) &= \frac{1}{n^2} \sum_{1 \leq i \neq j \leq n} \nabla_\beta \widehat{U}_i(t, d; \beta_0) \widehat{U}_j(t, d; \beta_0) \omega_{ij} \\ &\quad + \frac{\omega(0)}{n^2} \sum_{1 \leq i \leq n} \nabla_\beta \widehat{U}_i(t, d; \beta_0) \widehat{U}_i(t, d; \beta_0) \\ &= D_{n1}(t, d; \beta_0) + D_{n2}(t, d; \beta_0). \end{aligned}$$

Since $\widehat{I}_n(\beta_0)$ is the integral of $\widehat{I}_n(\cdot,\cdot;\beta_0)$ with respect to $\widehat{F}_{n,Y,\delta}$, the empirical distribution of the observations (Y,δ), the rate of the former could be derived from the uniform rate of the latter. In the following, when there is no danger of confusion, we simplify the writings and omit the arguments t, d, β_0. We will focus on the uniform rate of D_{n1} since the arguments for D_{n2}, which is uniformly negligible, are similar and much simpler. For this purpose we decompose $\widehat{U}_i$ and $\nabla_\beta \widehat{U}_i$ as follows:

$$\widehat{U}_i = \left\{\widehat{U}_i - \mathbb{E}\left[\widehat{U}_i \mid Y_i, \delta_i, X_i\right]\right\} + \left\{\mathbb{E}\left[\widehat{U}_i \mid Y_i, \delta_i, X_i\right] - U_i\right\} + U_i = V_{U,i} + B_{U,i} + U_i$$

and

$$\begin{aligned}\nabla_\beta \widehat{U}_i &= \left\{\nabla_\beta \widehat{U}_i - \mathbb{E}\left[\nabla_\beta \widehat{U}_i \mid X_i\right]\right\} + \left\{\mathbb{E}\left[\nabla_\beta \widehat{U}_i \mid X_i\right] - \mathbb{E}\left[\nabla_\beta U_i \mid X_i\right]\right\} + \mathbb{E}\left[\nabla_\beta U_i \mid X_i\right] \\ &= V_{\nabla,i} + B_{\nabla,i} + \mathbb{E}\left[\nabla_\beta U_i \mid X_i\right].\end{aligned}$$

By Lemma IX.1,

$$\sup_i \sup_{t,d} \left\{|B_{U,i}| + \|B_{\nabla,i}\|\right\} = O_{\mathbb{P}}(g^2),$$

and for any $\alpha \in (0,1)$,

$$\sup_i \sup_{t,d} \left\{|V_{U,i}| + g\,\|V_{\nabla,i}\|\right\} = O_{\mathbb{P}}(n^{-1/2} g^{\alpha/2-1}).$$

Then we can decompose

$$\begin{aligned}\frac{n-1}{n} D_{n1}(t,d;\beta_0) &= \frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} \mathbb{E}\left[\nabla_\beta U_i \mid X_i\right]\left\{U_j + V_{U,j}\right\}\omega_{ij} \\ &\quad + \frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} V_{\nabla,i}\{V_{U,j} + U_j\}\,\omega_{ij} \\ &\quad + O_{\mathbb{P}}(g^2) + O_{\mathbb{P}}\left(n^{-1/2} g^{\alpha/2}\right).\end{aligned}$$

By Lemma IX.2,

$$\sup_{t,d} \left\| \frac{1}{n(n-1)} \sum_{i\neq j} \mathbb{E}\left[\nabla_\beta U_i \mid X_i\right]\left\{U_j + V_{U,j}\right\}\omega_{ij} \right\| = O_{\mathbb{P}}\left(n^{-1/2}\right).$$

Moreover,

$$\sup_{t,d} \frac{1}{n(n-1)} \left\| \sum_{i\neq j} V_{\nabla,i}\{U_j + V_{U,j}\}\,\omega_{ij} \right\| = o_{\mathbb{P}}\left(n^{-1/2}\right).$$

Thus, integrating $D_{n1}(t,d;\beta_0)$ with respect to the empirical distribution function of the (Y_i,δ_i)'s yields the rate $O_{\mathbb{P}}(n^{-1/2})$. From similar arguments applied for $D_{n2}(t,d;\beta_0)$ one deuces the rate (VIII.4).

For the second order derivative we can decompose

$$\begin{aligned}\frac{1}{2}\nabla^2_{\beta\beta}\widehat{I}_n(t,d;\beta) &= \frac{1}{n^2}\sum_{1\le i,j\le n}\nabla_\beta\widehat{U}_i(t,d;\beta)\nabla_\beta\widehat{U}_j(t,d;\beta)^\top\omega_{ij}\\ &+ \frac{1}{n^2}\sum_{1\le i,j\le n}\nabla^2_{\beta\beta}\widehat{U}_i(t,d;\beta)\widehat{U}_j(t,d;\beta)\omega_{ij}\\ &= E_{n1}(t,d;\beta)+E_{n2}(t,d;\beta).\end{aligned}$$

It is shown in Lemma IX.4 in the Supplementary Material that, for any sequence $b_n\to 0$,

$$\sup_{t\in\mathbb{R},\ d\in\{0,1\}}\ \sup_{\|\beta-\beta_0\|\le b_n}|E_{n1}(t,d;\beta)-E_{n1}(t,d;\beta_0)| = o_{\mathbb{P}}(1)$$

and

$$\sup_{t\in\mathbb{R},\ d\in\{0,1\}}\ \sup_{\|\beta-\beta_0\|\le b_n}|E_{n2}(t,d;\beta)-E_{n2}(t,d;\beta_0)| = o_{\mathbb{P}}(1).$$

Here, the sequence (b_n) tending to zero is such that $\mathbb{P}[\|\widehat{\beta}-\beta_0\|\le b_n]\to 1$. On the other hand, by Lemmas IX.1 and IX.3 in the Supplementary Material, for any $\alpha\in(0,1)$,

$$\begin{aligned}E_{n2}(t,d;\beta_0) &= \frac{1}{n^2}\sum_{1\le i,j\le n}\left\{V_{\nabla^2,i}+B_{\nabla^2,i}+\mathbb{E}\left[\nabla^2_{\beta\beta}U_i\mid X_i\right]\right\}\left\{V_{U,j}+B_{U,j}+U_j\right\}\omega_{ij}\\ &= \frac{1}{n^2}\sum_{1\le i,j\le n}V_{\nabla^2,i}U_j\omega_{ij}+\frac{1}{n^2}\sum_{1\le i,j\le n}\left\{O_{\mathbb{P}}(n^{-1/2}g^{\alpha/2-3})\right\}\left\{O_{\mathbb{P}}(n^{-1/2}g^{\alpha/2-1})+O_{\mathbb{P}}(g^2)\right\}\omega_{ij}\\ &\quad+\frac{1}{n^2}\sum_{1\le i,j\le n}\left\{O_{\mathbb{P}}(g^2)\right\}\left\{O_{\mathbb{P}}(n^{-1/2}g^{\alpha/2-1})+O_{\mathbb{P}}(g^2)+U_j\right\}\omega_{ij}\\ &+\frac{1}{n^2}\sum_{1\le i,j\le n}\left\{O_{\mathbb{P}}(1)\right\}\left\{O_{\mathbb{P}}(n^{-1/2}g^{\alpha/2-1})+O_{\mathbb{P}}(g^2)\right\}\omega_{ij}+\frac{1}{n^2}\sum_{1\le i,j\le n}\mathbb{E}\left[\nabla^2_{\beta\beta}U_i\mid X_i\right]U_j\omega_{ij}.\end{aligned}$$

Given the uniform rates of the first and the last terms in the previous decomposition, see Lemma IX.3, the fact that α could be arbitrarily close to 1, and the fact that $ng^{3+a}\to\infty$ for some $a\in(0,1)$, deduce that

$$\begin{aligned}&\sup_{t\in\mathbb{R},\ d\in\{0,1\}}|E_{n2}(t,d;\beta_0)| = o_{\mathbb{P}}(1)\\ &\quad+\sup_{t\in\mathbb{R},\ d\in\{0,1\}}\left\{\left|\frac{1}{n^2}\sum_{1\le i,j\le n}V_{\nabla^2,i}U_j\omega_{ij}\right|+\left|\frac{1}{n^2}\sum_{1\le i,j\le n}\mathbb{E}\left[\nabla^2_{\beta\beta}U_i\mid X_i\right]U_j\omega_{ij}\right|\right\}\\ &\qquad= o_{\mathbb{P}}(1).\end{aligned}$$

It remains to investigate the uniform convergence of the term $E_{n1}(t,d;\beta_0)$. By Lemmas IX.1 and IX.2,

$$\sup_{t\in\mathbb{R},\ d\in\{0,1\}}\left|E_{n1}(t,d;\beta_0)-\mathbb{E}\left\{\mathbb{E}\left[\nabla_\beta U_1(t,d;\beta_0)|X_1\right]\mathbb{E}\left[\nabla_\beta U_2(t,d;\beta_0)^\top|X_2\right]\omega_{12}\right\}\right| = o_{\mathbb{P}}(1).$$

Thus

$$\int \left[E_{n1}(t,d;\beta_0) - \mathbb{E}\left[\mathbb{E}[\nabla_\beta U_1(t,d;\beta_0) \mid X_1]\mathbb{E}[\nabla_\beta U_2(t,d;\beta_0)^\top \mid X_2]\omega_{12}\right]\right] d\widehat{F}_{n,Y,\delta}(t,d) \to 0,$$

where $\widehat{F}_{n,Y,\delta}$ is the empirical distribution of the observations (Y_i,δ_i) with true distribution function $F_{Y,\delta}$. As a consequence, by the law of large numbers,

$$\int E_{n1}(t,d;\beta_0)d\widehat{F}_{n,Y,\delta}(t,d) - J(\beta_0) = o_{\mathbb{P}}(1),$$

where

$$J(\beta_0) = \int \mathbb{E}\left\{\mathbb{E}[\nabla_\beta U_1(t,d;\beta_0) \mid X_1]\mathbb{E}[\nabla_\beta U_2(t,d;\beta_0)^\top \mid X_2]\omega_{12}\right\} dF_{Y,\delta}(t,d). \qquad \text{(VIII.6)}$$

En explicit expression of $\mathbb{E}[\nabla_\beta U_i(t,d;\beta_0) \mid X_i]$ is provided in equation (IX.3) in the Supplementary Material. Our Assumption VIII.1-7 guarantees that $J(\beta_0)$ is positive definite. Gathering rates, deduce that the property (VIII.5) holds true and thus $\widehat{\beta}$ is $\sqrt{n}-$consistent.

For the asymptotic normality, by the definition of $\widehat{\beta}$ and a first order Taylor expansion,

$$0 = \nabla_\beta \widehat{I}_n(\widehat{\beta}) = \nabla_\beta \widehat{I}_n(\beta_0) + \nabla^2_{\beta\beta}\widehat{I}_n(\overline{\beta})(\widetilde{\widehat{\beta}} - \widetilde{\beta}_0),$$

where $\overline{\beta}$ is a vector between $\widehat{\beta}$ and β_0. Next, it suffices to follow the proof of the $\sqrt{n}-$asymptotic normality and deduce that

$$\nabla_\beta \widehat{I}_n(\beta_0) = 2\int \mathcal{V}_n(t,d;\beta_0)dF_{n,Y,\delta}(t,d) + o_{\mathbb{P}}(n^{-1/2}) \quad \text{and} \quad \nabla^2_{\beta\beta}\widehat{I}_n(\overline{\beta}) = J(\beta_0) + o_{\mathbb{P}}(1),$$

for any sequence $\overline{\beta}$ between $\widehat{\beta}$ and β_0, and

$$\mathcal{V}_n(t,d;\beta_0) = \frac{1}{n(n-1)}\sum_{i\neq j}\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0) \mid X_i\right] U_j(t,d;\beta_0)\,\omega_{ij}.$$

One can decompose

$$\mathcal{V}_n(t,d;\beta_0) = \frac{1}{n}\sum_{1\leq j\leq n}\mathbb{E}\left\{\mathbb{E}\left[\nabla_\beta U(t,d;\beta_0) \mid X\right]\omega(X-X_j) \mid X_j\right\} U_j(t,d;\beta_0)$$
$$+ \frac{1}{n(n-1)}\sum_{i\neq j}\left[\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0) \mid X_i\right]\omega_{ij} - \mathbb{E}\left\{\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0) \mid X_i\right]\omega_{ij} \mid X_j\right\}\right] U_j(t,d;\beta_0)$$
$$= \mathcal{V}_{1n}(t,d;\beta_0) + \mathcal{V}_{2n}(t,d;\beta_0).$$

The maximal inequality (IX.2) implies that the degenerate second order $U-$process $\mathcal{V}_{2n}$ is of uniform rate $O_{\mathbb{P}}(n^{-1})$, and hence negligible. On the other hand,

$$\int \mathcal{V}_{1n}(t,d;\beta_0)dF_{n,Y,\delta}(t,d) = \int \mathcal{V}_{1n}(t,d;\beta_0)dF_{Y,\delta}(t,d) + o_{\mathbb{P}}(n^{-1/2}),$$

as shown in Lemma IX.7 in the Supplementary Material. Finally, the multivariate Central Limit Theorem implies that $\int \mathcal{V}_{1n}(t,d;\beta_0)dF_{Y,\delta}(t,d)$ is asymptotically normal and thus

$$\sqrt{n}\nabla_\beta \widehat{I}_n(\beta_0) \rightsquigarrow N(0,\Sigma(\beta_0)),$$

where

$$\Sigma(\beta_0) = 4\mathbb{E}\left[\psi(Y,\delta,X;\beta_0)\psi(Y,\delta,X;\beta_0)^\top\right], \tag{VIII.7}$$

is a positive definite $(p-1)\times(p-1)-$matrix and

$$\psi(Y_j,\delta_j,X_j;\beta_0) = \int \mathbb{E}\left\{\mathbb{E}\left[\nabla_\beta U(t,d;\beta_0) \mid X\right]\omega(X-X_j) \mid X_j\right\} U_j(t,d;\beta_0)dF_{Y,\delta}(t,d).$$

Finally deduce

$$\sqrt{n}(\widehat{\beta}-\beta_0) \rightsquigarrow N(0,V(\beta_0)) \qquad \text{where} \qquad V(\beta_0) = J(\beta_0)^{-1}\Sigma(\beta_0)J(\beta_0)^{-1},$$

and $J(\beta_0)$ and $\Sigma(\beta_0)$ are defined as in equations (VIII.6) and (VIII.7). Now the proof of Proposition 4.1 is complete. ■

VIII.4 Proof of Proposition 5.1

The idea of the proof is to show that

$$\sup_{x\in\overline{\mathcal{X}}}\sup_{t\leq\tau}\left|\widehat{\Lambda}_{T,\widehat{\beta}}((-\infty,t] \mid x^\top\widehat{\beta}) - \widehat{\Lambda}_{T,\beta_0}((-\infty,t] \mid x^\top\beta_0)\right| = o_{\mathbb{P}}(n^{-1/2}h^{-1/2}), \tag{VIII.8}$$

and then to apply the i.i.d. representations of Du & Akritas (2002) and Lopez (2011) for $\widehat{\Lambda}_{T,\beta_0}((-\infty,t] \mid x^\top\beta_0)$. The representation for $\widehat{F}_{T,\beta_0}((-\infty,t] \mid x^\top\beta_0)$ will follow easily. Since for any given β

$$\widehat{\Lambda}_{T,\beta}((-\infty,t] \mid x^\top\beta) = \frac{1}{nh}\sum_{i=1}^n \frac{\mathbf{1}\{Y_i\leq t\}\delta_i K((X_i-x)^\top\beta/h)}{\frac{1}{nh}\sum_{k=1}^n \mathbf{1}\{Y_k\geq Y_i\}K((X_k-x)^\top\beta/h)},$$

it suffices, on one hand, to show that

$$\sup_{\|\beta-\beta_0\|\leq b_n}\sup_{t\leq\tau}\sup_{x\in\overline{\mathcal{X}}}|\Delta_n(t,x;\beta)| = o_{\mathbb{P}}(n^{-1/2}h^{-1/2}),$$

where

$$\Delta_n(t,x;\beta) = \frac{1}{nh}\sum_{i=1}^n \xi(Y_i,\delta_i;t)\left[K((X_i-x)^\top\beta/h) - K((X_i-x)^\top\beta_0/h)\right],$$

with

$$\xi(Y_i,\delta_i;t) = \mathbf{1}\{Y_i\leq t\}\delta_i \quad \text{or} \quad \xi(Y_i,\delta_i;t) = \mathbf{1}\{Y_i\geq t\}, \qquad t\leq\tau,$$

and $b_n = O_{\mathbb{P}}(n^{-1/2})$. On the other hand, to guarantee that, for some constant $c > 0$,

$$D_n = \inf_{x\in\overline{\mathcal{X}}} \frac{1}{nh}\sum_{k=1}^{n} \mathbf{1}\{Y_k \geq \tau\} K((X_k - x)^\top \beta_0/h) \geq c$$

with probability tending to 1.

To uniformly bound Δ_n let us write

$$\begin{aligned}|\Delta_n(t, x; \beta)| &\leq |\Delta_{1n}(t, x; \beta) - \Delta_{1n}(t, x; \beta_0)| \\ &\quad + \left|\mathbb{E}\left[\xi(Y, \delta; t)h^{-1}\left\{K((X - x)^\top \beta/h) - K((X - x)^\top \beta_0/h)\right\}\right]\right| \\ &= |\Delta_{1n}(t, x; \beta) - \Delta_{1n}(t, x; \beta_0)| + |\Delta_{2n}(t, x; \beta, \beta_0)|,\end{aligned}$$

where

$$\Delta_{1n}(t, x; \beta) = \frac{1}{nh}\sum_{i=1}^{n}\left\{\xi(Y_i, \delta_i; t)K((X_i - x)^\top \beta/h) - \mathbb{E}\left[\xi(Y, \delta; t)K((X - x)^\top \beta/h)\right]\right\}.$$

For any fixed β, the process $\Delta_{1n}(t, x; \beta)$ indexed by $t \leq \tau$ and $x \in \overline{\mathcal{X}}$ is of uniform rate $O_{\mathbb{P}}(n^{-1/2}h^{-1/2}\log^{1/2} n)$. Letting β to approach β_0 at the rate $O_{\mathbb{P}}(n^{-1/2})$, we could use the modulus of continuity of the process $\Delta_{1n}(t, x; \beta)$ indexed by t, x and β and thus derive the required uniform rate $o_{\mathbb{P}}(n^{-1/2}h^{-1/2})$ for $\Delta_{1n}(t, x; \beta) - \Delta_{1n}(t, x; \beta_0)$. By Taylor expansion, suitable changes of variable and the assumptions guaranteeing that $|f_\beta(\cdot) - f_{\beta_0}(\cdot)|/\|\beta - \beta_0\|$ is bounded and $f_{\beta_0}(\cdot)$ is Lipschitz continuous, $\Delta_{2n}(t, x; \beta, \beta_0)$ could be also shown to be of uniform rate $o_{\mathbb{P}}(n^{-1/2}h^{-1/2})$. The details are provided in Lemma IX.6 in the Supplementary Material.

To bound D_n, let us decompose

$$\begin{aligned}D_n &= \inf_{x\in\overline{\mathcal{X}}} \mathbb{E}\{H_{\beta_0}([\tau, \infty) \mid X^\top \beta_0)h^{-1}K((X - x)^\top \beta_0/h)\} \\ &\quad - \sup_{x\in\overline{\mathcal{X}}} \frac{1}{nh}\sum_{k=1}^{n}\left[\mathbf{1}\{Y_k \geq \tau\}K((X_k - x)^\top\beta_0/h) - \mathbb{E}\{H_{\beta_0}([\tau, \infty) \mid X^\top\beta_0)K((X - x)^\top \beta_0/h)\}\right] \\ &= D_{1n} - D_{2n}.\end{aligned}$$

By standard results in the uniform convergence of kernel estimators, see for instance Theorem 4 of Einmahl & Mason (2005), $D_{2n} = O_{\mathbb{P}}(n^{-1/2}h^{-1/2}\log^{1/2} n) = o_{\mathbb{P}}(1)$. (A slightly slower uniform rate, but still negligible, could be obtained using the Maximal Inequality of Sherman (1994) recalled in the Supplementary Material.) It remains to show that D_{1n} stays away from zero. By a change of variables and Taylor expansion,

$$\mathbb{E}\{H_{\beta_0}([\tau, \infty) \mid X^\top \beta_0)h^{-1}K((X - x)^\top \beta_0/h)\} = H_{\beta_0}([\tau, \infty) \mid x^\top \beta_0)f_{\beta_0}(x^\top \beta_0) + O_{\mathbb{P}}(h^2),$$

uniformly with respect to $x \in \overline{\mathcal{X}}$. By construction and assumptions,

$$\inf_{x\in\overline{\mathcal{X}}} H_{\beta_0}([\tau, \infty) \mid x^\top \beta_0)f_{\beta_0}(x^\top \beta_0) > 0,$$

so that D_n is uniformly bounded away from zero. Gathering facts, one deduces the uniform rate in equation (VIII.8). Now, by the Duhamel identity, see Gill & Johansen (1990), page 1635, one deduces

$$\sup_{x\in\overline{\mathcal{X}}}\sup_{t\leq\tau}\left|\widehat{F}_{T,\widehat{\beta}}((-\infty,t]\mid x^\top\widehat{\beta})-\widehat{F}_{T,\beta_0}((-\infty,t]\mid x^\top\beta_0)\right|=o_{\mathbb{P}}(n^{-1/2}h^{-1/2}).$$

See also Xu & Peng (2014), page 14. Then the representation for $\widehat{F}_{T,\widehat{\beta}}$ follows from the i.i.d. representation of the conditional survival function derived by Du & Akritas (2002) and Lopez (2011). Now the proof is complete.

VIII.5 Proof of Corollary 5.2

Let $\tau=\sup_{x\in\mathcal{X}}\tau_{H_1}(x;\beta_0)$. By definition $Y^1_{(n)}\leq\tau$ and

$$\widehat{F}_{T,\widehat{\beta}}((Y^1_{(n)},\infty])\mid x^\top\widehat{\beta})=\widehat{F}_{T,\widehat{\beta}}((\tau,\infty])\mid x^\top\widehat{\beta}),\qquad\forall x\in\mathcal{X}.$$

Next, by assumption we have $\tau<\inf_{x\in\overline{\mathcal{X}}}\tau_H(x;\beta_0)$. Since obviously

$$\pi(x)=F_{T,\beta_0}((\tau_H(x),\infty]\mid x^\top\beta_0),\qquad\forall x\in\mathcal{X},$$

the $\sqrt{nh}-$asymptotic normality follows directly from the i.i.d. representation (V.1) and a central limit theorem applied for the i.i.d. sum $n^{-1}\sum_{i=1}^n\eta_{F_T,i}(\tau,x^\top\beta_0)$ considered with some $x\in\overline{\mathcal{X}}$.

IX Supplementary material: comments, technical lemmas and proofs

In the following $C, C_1, C_2, C', \cdots$ represent constants, independent of the sample size. Their value may change from line to line. For nonnegative quantities a_n and r_n possibly depending on n, we will use the notation $a_n \lesssim r_n$ to indicate that there exists some constant C such that for each n, $a_n \leq C r_n$.

IX.1 Comments on the Assumptions VIII.1

The exponential moment imposed by condition 1 implies that the largest value of the sample $\|X_1\|, \cdots, \|X_n\|$ has the rate $O_{\mathbb{P}}(\log n)$. However, X is not required to be bounded or to have a density. The condition 2 guarantees that $F_{C,\beta}(\mathbb{R} \mid z) = 1$, $\forall z \in \mathbb{R}$, $\forall \beta \in \mathcal{B}$. The condition 4 is a classical condition used to prove consistency. The regularity imposed by condition 6, together with the Taylor expansion, serve to show that biases are negligible. In order to have a nondegenerate limit for $\widehat{\beta}$ we need the matrix

$$\int \mathbb{E}\left[\mathbb{E}[\nabla_\beta U_1(t,d;\beta_0) \mid X_1]\mathbb{E}[\nabla_\beta U_2(t,d;\beta_0)^\top \mid X_2]\omega_{12}\right] dF_{Y,\delta}(t,d) \tag{IX.1}$$

to be definite positive. This means, if $v \in \mathbb{R}^{d-1}$ satisfies

$$\int \mathbb{E}\left[v^\top\mathbb{E}[\nabla_\beta U_1(t,d;\beta_0) \mid X_1]\mathbb{E}[\nabla_\beta U_2(t,d;\beta_0)^\top \mid X_2]v\ \omega_{12}\right] dF_{Y,\delta}(t,d) = 0,$$

then necessarily $v = 0$. By construction, the integrand in the last display is nonnegative. For any t, d such that the integrand is positive, by the Inverse Fourier Transform and the fact that $\mathcal{F}[\omega] > 0$, deduce that $\mathbb{E}[v^\top\nabla_\beta U(t,d;\beta_0) \mid X] = 0$ almost surely. This means that the variance of $\mathbb{E}[v^\top\nabla_\beta U(t,d;\beta_0) \mid X]$ is zero. Then, the condition of positive definiteness of the variance defined in (VIII.3) for a set of values (t,d) of positive probability implies that necessarily $v = 0$. Condition 8 is used to control the difference of the second order derivative $\nabla^2_{\beta\beta}\widehat{I}_n(\beta) - \nabla^2_{\beta\beta}\widehat{I}_n(\beta_0)$ when $\beta - \beta_0$ tends to zero. The properties of the kernel $L(\cdot)$ allow for the Taylor expansion used to study the bias of various quantities and to guarantee the VC-class property for the families indexing various $U-$processes appearing in the proof. (The definition of VC-class is recalled in the Supplementary Material.) The bandwidth g should be small enough in order to make bias terms negligible. This explains the condition $ng^4 \to 0$. The other condition of g is a convenient restriction to control variance terms.

IX.2 Technical lemmas and proofs

Let $\mathcal{S}$ be an arbitrary space where the i.i.d. observations take value. Let m be a positive integer and $\mathcal{F}$ a class of real-valued functions on the product space $\mathcal{S}^m = \mathcal{S} \otimes \cdots \otimes \mathcal{S}$. Let F be an envelope for $\mathcal{F}$, that is $\sup_{\mathcal{F}} |f(\cdot)| \leq F(\cdot)$. For ν a probability measure on $\mathcal{S}^m$

and $\varepsilon > 0$, let $N(\varepsilon \|F\|_{L^2(\nu)}, \mathcal{F}, L^2(\nu))$, denote the covering number, that is the minimal number of balls of radius $\varepsilon \|F\|_{L^2(\nu)}$ in $L^2(\nu)$ needed to cover $\mathcal{F}$. See van der Vaart and Wellner (1996) or Sherman (1994) for the definitions. To derive our main results, we will need to derive the rates of uniform convergence for degenerate $U-$processes of order m indexed by VC (or Euclidian) classes of functions $\mathcal{F}$. A VC-class with constants (A, V) is a class of functions with covering number bounded by the polynomial $A\varepsilon^{-V}$, $\forall 0 < \varepsilon \leq 1$. A $U-$processes of order m indexed by $\mathcal{F}$, denoted by $U_n^m f$, is degenerate if for each $f \in \mathcal{F}$, $\int_{\mathcal{S}} f(s_1, \ldots, s_{j-1}, \cdot, s_{j+1}, \ldots, s_m) dP = 0$, $j = 1, \ldots, m$, where P is the probability distribution of the observations. The case $m = 1$ corresponds to empirical processes. Let us recall a simplified version of a general result of Sherman (1994) that could apply in all cases we need to investigate below. It is certainly not the sharpest result of this kind but, in particular, it allows to work with squared integrable envelopes so that is no longer necessary to truncate the functions from $\mathcal{F}$ and treat the tails separately. Hence, for the purposes of our theoretical study, we consider the uniform bound below a readable compromise that allows for only little loss of generality in the conditions on the bandwidth g. For the classes of functions we consider, the VC-class property is a direct consequence of standard results available, for instance, in Nolan and Pollard (1987), Pakes and Pollard (1989), Sherman (1994), van der Vaart and Wellner (1996). Hence, in following we will omit the details for justifying the VC-class property.

Maximal Inequality [Sherman (1994), Main Corollary] *Let $\mathcal{F}$ be a class of degenerate functions on $\mathcal{F}$ on $\mathcal{S}^m$, $m \geq 1$. Suppose $\mathcal{F}$ is a VC-class of parameters (A, V) for a squared integrable envelope F. Then, for any $\alpha \in (0, 1)$,*

$$\mathbb{E} \sup_{\mathcal{F}} \left| n^{m/2} U_n^p f \right| \leq \Lambda \left[\mathbb{E} \sup_{\mathcal{F}} \left(U_{2n}^m f^2 \right)^\alpha \right]^{1/2} \tag{IX.2}$$

with $\Lambda > 0$ a constant depending on A, V, m and α, but independent of n.

Lemma IX.1 *Under the condition of Proposition 4.1:*

1.
$$\sup_i \sup_{t,d} \{ |B_{U,i}| + \|B_{\nabla,i}\| \} = O_{\mathbb{P}}(g^2).$$

2. For any $\alpha \in (0, 1)$,

$$\sup_i \sup_{t,d} \{ |V_{U,i}| + g \|V_{\nabla,i}\| \} = O_{\mathbb{P}}(n^{-1/2} g^{\alpha/2 - 1}).$$

Proof of Lemma IX.1. 1. For the uniform rate on the bias $B_{U,i}$ it suffices to show that

$$\mathbb{E}\left[g^{-1} L((X - X_i)^\top \beta_0 / g) \mid X_i \right] - f_{\beta_0}(X_i^\top \beta_0) = O_{\mathbb{P}}(g^2)$$

and

$$\mathbb{E}\left[H_{d,\beta_0}((-\infty,t]\,|\,X^\top\beta_0)g^{-1}L((X-X_i)^\top\beta_0/g)\,|\,X_i\right] - H_{d,\beta_0}((-\infty,t]\,|\,X_i^\top\beta_0)f_\beta(X_i^\top\beta_0) = O_{\mathbb{P}}(g^2),$$

uniformly with respect to i, t and d. These uniform rates follow by standard change of variables and Taylor expansion of the functions $z \to f_{\beta_0}(z)$ and $z \to H_{d,\beta}((-\infty,t] \mid z)f_{\beta_0}(z)$, which, by our assumptions, have the required regularity.

For the uniform rate on the bias $B_{\nabla^2,i}$ let us write

$$\begin{aligned}\mathbb{E}\left[\nabla_\beta\widehat{U}_i(t,d;\beta_0)\mid X_i\right] &= H_{d,\beta_0}((-\infty,t]\mid X_i^\top\beta_0)\\ &\quad\times\mathbb{E}\left[\left(\widetilde{X}_i-\mathbb{E}[\widetilde{X}_k\mid X_k^\top\beta_0]\right)g^{-2}L'((X_i-X_k)^\top\beta_0/g)\mid X_i\right]\\ &- \mathbb{E}\left[H_{d,\beta_0}((-\infty,t]\mid X_k^\top\beta_0)\left(\widetilde{X}_i-\mathbb{E}[\widetilde{X}_k\mid X_k^\top\beta_0]\right)\right.\\ &\quad\left.\times\, g^{-2}L'((X_i-X_k)^\top\beta_0/g)\mid X_i\right].\end{aligned}$$

We used the fact that for each t and d, by the single-index assumption on the law of (Y,δ), we have $H_d((-\infty,t]\mid X_i) = H_{d,\beta_0}((-\infty,t]\mid X_i^\top\beta_0)$ and

$$\mathbb{E}\left[\widetilde{X}_k\mathbf{1}\{Y_k\le t,\delta_k=d\}\mid X_k^\top\beta_0\right] = H_{d,\beta_0}((-\infty,t]\mid X_k^\top\beta_0)\mathbb{E}[\widetilde{X}_k\mid X_k^\top\beta_0].$$

(Recall that $\widetilde{X}_i\in\mathbb{R}^{p-1}$ denotes the $(p-1)-$dimension vector of the last components of X_i.) Next, by integration by parts and Taylor expansion

$$\begin{aligned}&\mathbb{E}\left[\left(\widetilde{X}_i-\mathbb{E}[\widetilde{X}_k\mid X_k^\top\beta_0]\right)g^{-2}L'((X_i-X_k)^\top\beta_0/g)\mid X_i\right]\\ &\quad=\int\partial_z\left(\left(\widetilde{X}_i-\mathbb{E}[\widetilde{X}\mid X^\top\beta_0=\cdot]\right)f_{\beta_0}(\cdot)\right)(z)g^{-1}L((X_i^\top\beta_0-z)/g)dz\\ &\quad\quad=\partial_z\left(\left(\widetilde{X}_i-\mathbb{E}[\widetilde{X}\mid X^\top\beta_0=\cdot]\right)f_{\beta_0}(\cdot)\right)(X_i^\top\beta_0)+O_{\mathbb{P}}(g^2),\end{aligned}$$

where for any univariate map $z\mapsto\zeta(z)$, $\partial_z\zeta$ denotes its derivative. Similarly,

$$\begin{aligned}&\mathbb{E}\left[H_{d,\beta_0}((-\infty,t]\mid\cdot)\left(\widetilde{X}_i-\mathbb{E}[\widetilde{X}_k\mid X_k^\top\beta_0]\right)g^{-2}L'((X_i-X_k)^\top\beta_0/g)\mid X_i\right]\\ &\quad=\int\partial_z\left(H_{d,\beta_0}((-\infty,t]\mid\cdot)\left(\widetilde{X}_i-\mathbb{E}[\widetilde{X}\mid X^\top\beta_0=\cdot]\right)f_{\beta_0}(\cdot)\right)(z)g^{-1}L((X_i^\top\beta_0-z)/g)dz\\ &\quad\quad=\partial_z\left(H_{d,\beta_0}((-\infty,t]\mid\cdot)\left(\widetilde{X}_i-\mathbb{E}[\widetilde{X}\mid X^\top\beta_0=\cdot]\right)f_{\beta_0}(\cdot)\right)(X_i^\top\beta_0)+O_{\mathbb{P}}(g^2).\end{aligned}$$

By our assumptions, the last two $O_{\mathbb{P}}(g^2)$ rates above are uniform with respect to i, t and d. Deduce that

$$\sup_i\sup_{t,d}\left\|\mathbb{E}\left[\nabla_\beta\widehat{U}_i(t,d;\beta_0)\mid X_i\right]-\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0)\mid X_i\right]\right\| = O_{\mathbb{P}}(g^2),$$

where

$$\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0) \mid X_i\right] = -\left(\widetilde{X}_i - \mathbb{E}[\widetilde{X}_i \mid X_i^\top \beta_0]\right) \\ \times f_{\beta_0}(X_i^\top \beta_0)\, \partial_z \left(H_{d,\beta_0}((-\infty, t] \mid \cdot)\right)(X_i^\top \beta_0). \quad \text{(IX.3)}$$

One could note that $\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0)\right] = 0$.

2. From the definitions we obtain

$$\begin{aligned}
&\widehat{U}_i(t,d;\beta_0) - \mathbb{E}\left[\widehat{U}_i(t,d;\beta_0) \mid Y_i, \delta_i, X_i\right] \\
&= \mathbf{1}\{Y_i \le t, \delta_i = d\} \frac{1}{n} \sum_{k=1}^n \left\{ g^{-1} L((X_i - X_k)^\top \beta_0 / g) - \mathbb{E}\left[g^{-1} L((X_i - X_k)^\top \beta_0 / g) \mid X_i \right] \right\} \\
&\quad + \frac{1}{n} \sum_{k=1}^n \left\{ \mathbf{1}\{Y_k \le t, \delta_k = d\} g^{-1} L((X_i - X_k)^\top \beta_0 / g) \right. \\
&\qquad \left. - \mathbb{E}\left[\mathbf{1}\{Y_k \le t, \delta_k = d\} g^{-1} L((X_i - X_k)^\top \beta_0 / g) \mid X_i \right] \right\}.
\end{aligned}$$

The uniform rate follows by applying twice the Maximal Inequality (IX.2) with $m = 1$ and α close to 1. Let us detail one these situations, the other one is similar. For any $w \in \mathbb{R}^p$, $t \ge 0$ and $d \in \{0,1\}$, let

$$f = \mathbf{1}\{\cdot \le t, \cdot = d\} L((w - \cdot)^\top \beta_0 / g) - \mathbb{E}\left[\mathbf{1}\{\cdot \le t, \cdot = d\} L((w - \cdot)^\top \beta_0 / g) \right]$$

a function of the variables Y, δ, X that depends on w, t, d. By the imposed assumptions, the family of such functions f indexed by w, t, d is Euclidean for a bounded envelope. Moreover, since for any real numbers a and b, $(a-b)^2 \le 2(a^2 + b^2)$, and L is bounded,

$$f^2 \le 2\{ L((w - \cdot)^\top \beta_0 / g) + \mathbb{E}\left[L((w - \cdot)^\top \beta_0 / g) \right] \}.$$

One can apply inequality (IX.2) with $m = 1$. It remains to bound the expectation on the right hand side. For this, up to a constant, one can use the bound

$$\mathbb{E}\left[L((X_1 - X_2)^\top \beta_0 / g) \right] \lesssim g$$

and Jensen's inequality to deduce that the right hand side of the inequality (IX.2) is of rate $g^{\alpha/2}$. The uniform rate of $V_{U,i}$ follows immediately. Modulo minor modifications and using also the assumption (VIII.2) and Cauchy-Schwarz inequality, the uniform rate of $V_{\Delta,i}$ could be obtained by the same arguments. ■

Lemma IX.2 *Under the condition of Proposition 4.1:*

1.

$$\sup_{t,d} \left\| \frac{1}{n(n-1)} \sum_{i \ne j} \mathbb{E}[\nabla_\beta U_i(t,d;\beta_0) \mid X_i] \mathbb{E}[\nabla_\beta U_j(t,d;\beta_0) \mid X_j]^\top \omega_{ij} \right. \\ \left. - \mathbb{E}\left[\mathbb{E}[\nabla_\beta U_1(t,d;\beta_0) \mid X_1] \mathbb{E}[\nabla_\beta U_2(t,d;\beta_0) \mid X_2]^\top \omega_{12} \right] \right\| = o_{\mathbb{P}}(1).$$

2.

$$\sup_{t,d}\left\|\frac{1}{n(n-1)}\sum_{i\neq j}\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0)\mid X_i\right]\{U_j+V_{U,j}\}\,\omega_{ij}\right\| = O_{\mathbb{P}}\left(n^{-1/2}\right).$$

3.

$$\sup_{t,d}\left\|\frac{1}{n(n-1)}\sum_{1\le i\neq j\le n}V_{\nabla,i}\{U_j+V_{U,j}\}\,\omega_{ij}\right\| = o_{\mathbb{P}}\left(n^{-1/2}\right).$$

Proof of Lemma IX.2. 1. Use Hoeffding decomposition in degenerate $U-$statistics. Next use twice the maximal inequality (IX.2) or Corollary 4 of Sherman (1994).

2. The assumptions guarantee the VC-class property for the class of functions

$$z\mapsto \partial_z\left(H_{d,\beta_0}((-\infty,t]\mid\cdot)\right)(z),\qquad t\in(-\infty,\infty),$$

for $d=0$ and $d=1$. The VC-class property for the class $\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0)\mid X_i\right]U_j\;\omega_{ij}$, indexed by t and d, follows. Since

$$\mathbb{E}\left[\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0)\mid X_i\right]U_j\;\omega_{ij}\right]=\mathbb{E}\left[\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0)\mid X_i\right]\mathbb{E}[U_j(t,d;\beta_0)\mid X_j]\;\omega_{ij}\right]=0,$$

Corollary 4 of Sherman (1994) provides the uniform rate $O_{\mathbb{P}}\left(n^{-1/2}\right)$ for the $U-$process

$$\mathcal{V}_n(t,d;\beta_0)=\frac{1}{n(n-1)}\sum_{i\neq j}\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0)\mid X_i\right]U_j(t,d;\beta_0)\;\omega_{ij}.$$

For the remaining part in the $U-$process investigated at this point, simplifying the notation $\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0)\mid X_i\right]$ to $\mathbb{E}\left[\nabla_\beta U_i\mid X_i\right]$, we can write

$$\begin{aligned}
\frac{1}{n(n-1)}\sum_{i\neq j}\mathbb{E}\left[\nabla_\beta U_i\mid X_i\right]V_{U,j}\;\omega_{ij} &= \frac{g^{-1}}{n(n-1)}\sum_{i\neq j\neq k}\mathbb{E}\left[\nabla_\beta U_i\mid X_i\right]\\
&\times\left\{\mathbf{1}\{Y_j\le t,\delta_j=d\}[L((X_j-X_k)^\top\beta_0/g)-\mathbb{E}(L((X_j-X)^\top\beta_0/g)\mid X_j)]\right\}\;\omega_{ij}\\
&-\frac{g^{-1}}{n(n-1)}\sum_{i\neq j\neq k}\mathbb{E}\left[\nabla_\beta U_i\mid X_i\right]\left\{[\mathbf{1}\{Y_k\le t,\delta_k=d\}L((X_j-X_k)^\top\beta_0/g)\right.\\
&\left.-\mathbb{E}(\mathbf{1}\{Y\le t,\delta=d\}L((X_j-X)^\top\beta_0/g)\mid X_j)]\right\}\;\omega_{ij}\\
&+\text{diagonal terms of smaller order.}\\
&=g^{-1}[\mathcal{U}_{1n}(t,d;\beta_0)-\mathcal{U}_{2n}(t,d;\beta_0)]+\text{diagonal terms of smaller order.}
\end{aligned}$$

Next, we have to apply the Hoeffding decomposition to the third order $U-$processes $\mathcal{U}_{1n}(t,d;\beta_0)$ and $\mathcal{U}_{2n}(t,d;\beta_0)$ indexed by t and d. By the Maximal Inequality (IX.2), the uniform rate of the degenerate third order $U-$processes in the Hoeffding decomposition of $\mathcal{U}_{1n}(t,d,\beta_0)$ and $\mathcal{U}_{2n}(t,d;\beta_0)$ is $O_{\mathbb{P}}(g^{\alpha/2}n^{-3/2})$. This rate, divided by g, is negligible compared to $O_{\mathbb{P}}(n^{-1/2})$. Now, let us denote $\upsilon_{1\mid i,j,k}$ and $\upsilon_{2\mid i,j,k}$ the functions defining $\mathcal{U}_{1n}(t,d;\beta_0)$ and $\mathcal{U}_{2n}(t,d;\beta_0)$. Moreover, for any $i\neq j$, let us denote

$$\mathbb{E}_{ij}[\cdots] = \mathbb{E}[\cdots \mid Y_i,\delta_i,X_i,Y_j,\delta_j,X_j] \quad \text{and} \quad \mathbb{E}_i[\cdots] = \mathbb{E}[\cdots \mid Y_i,\delta_i,X_i].$$

By construction, for each t and d,

$$\mathbb{E}_{ij}[\upsilon_{1\mid i,j,k}(t,d;\beta_0)] = \mathbb{E}_{ij}[\upsilon_{2\mid i,j,k}(t,d;\beta_0)] = 0. \tag{IX.4}$$

Next, by the maximal inequality of Sherman (1994), the two remaining degenerate second order $U-$processes in the Hoeffding decomposition of $\mathcal{U}_{1n}(t,d;\beta_0)$ and $\mathcal{U}_{2n}(t,d;\beta_0)$ are of uniform rate is $O_{\mathbb{P}}(g^{\alpha/2}n^{-1})$. This rate, divided by g, is still negligible compared to $O_{\mathbb{P}}(n^{-1/2})$. Finally, the property (IX.4) implies

$$\mathbb{E}_i[\upsilon_{l\mid i,j,k}(t,d;\beta_0)] = \mathbb{E}_j[\upsilon_{l\mid i,j,k}(t,d;\beta_0)] = 0, \quad \text{for} \quad l=1 \quad \text{and} \quad l=2,$$

so that it remains to investigate only the empirical processes defined by $\mathbb{E}_k[\upsilon_{1\mid i,j,k}(t,d;\beta_0)]$ and $\mathbb{E}_k[\upsilon_{2\mid i,j,k}(t,d;\beta_0)]$. By elementary calculations, uniformly with respect to t and d,

$$\begin{aligned}\mathbb{E}_k[\upsilon_{1\mid i,j,k}(t,d;\beta_0)] &= \mathbb{E}_k\left[\mathbb{E}[\nabla_\beta U_i \mid X_i] H_{d,\beta_0}((-\infty,t] \mid X_j^\top\beta_0) L((X_j-X_k)^\top\beta_0/g)\,\omega_{ij}\right]\\ &\quad - \mathbb{E}\left[\mathbb{E}[\nabla_\beta U_i \mid X_i] H_{d,\beta_0}((-\infty,t] \mid X_j^\top\beta_0)\mathbb{E}[L((X_j-X)^\top\beta_0/g) \mid X_j]\,\omega_{ij}\right]\\ &= g\mathbb{E}\left[\mathbb{E}[\nabla_\beta U_i \mid X_i]\omega_{ik} \mid X_k^\top\beta_0\right] H_{d,\beta_0}((-\infty,t] \mid X_k^\top\beta_0) f_{\beta_0}(X_k^\top\beta_0)\\ &\quad - g\mathbb{E}\left\{\mathbb{E}\left[\mathbb{E}[\nabla_\beta U_i \mid X_i]\omega_{ij} \mid X_j^\top\beta_0\right] H_{d,\beta_0}((-\infty,t] \mid X_j^\top\beta_0) f_{\beta_0}(X_j^\top\beta_0)\right\} + O_{\mathbb{P}}(g^3).\end{aligned}$$

On the other hand, uniformly with respect to t and d,

$$\begin{aligned}\mathbb{E}_k[\upsilon_{2\mid i,j,k}(t,d;\beta_0)] &= \mathbb{E}_k\left[\mathbb{E}[\nabla_\beta U_i \mid X_i] H_{d,\beta_0}((-\infty,t] \mid X_k^\top\beta_0) L((X_j-X_k)^\top\beta_0/g)\,\omega_{ij}\right]\\ &\quad - \mathbb{E}\left[\mathbb{E}[\nabla_\beta U_i \mid X_i]\mathbb{E}[H_{d,\beta_0}((-\infty,t] \mid X^\top\beta_0) L((X_j-X)^\top\beta_0/g) \mid X_j]\,\omega_{ij}\right]\\ &= g\mathbb{E}\left[\mathbb{E}[\nabla_\beta U_i \mid X_i]\omega_{ik} \mid X_k^\top\beta_0\right] H_{d,\beta_0}((-\infty,t] \mid X_k^\top\beta_0) f_{\beta_0}(X_k^\top\beta_0)\\ &\quad - g\mathbb{E}\left\{\mathbb{E}\left[\mathbb{E}[\nabla_\beta U_i \mid X_i]\omega_{ij} \mid X_j^\top\beta_0\right] H_{d,\beta_0}((-\infty,t] \mid X_j^\top\beta_0) f_{\beta_0}(X_j^\top\beta_0)\right\} + O_{\mathbb{P}}(g^3).\end{aligned}$$

As a consequence, uniformly with respect to t and d,

$$g^{-1}\left\{\mathbb{E}_k[\upsilon_{1\mid i,j,k}(t,d;\beta_0)] - \mathbb{E}_k[\upsilon_{2\mid i,j,k}(t,d;\beta_0)]\right\} = O_{\mathbb{P}}(g^2) = o_{\mathbb{P}}(n^{-1/2}).$$

Deduce that

$$\sup_{t,d}\left\|\frac{1}{n(n-1)}\sum_{i\neq j}\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0) \mid X_i\right] V_{U,j}\,\omega_{ij}\right\| = o_{\mathbb{P}}\left(n^{-1/2}\right). \tag{IX.5}$$

3. The same type of arguments as those used to derive the uniform rate (IX.5) apply. The details are omitted. ∎

Let us introduce some more notation:

$$B_{\nabla^2,i} = \mathbb{E}\left[\nabla^2_{\beta\beta}\widehat{U}_i \mid X_i\right] - \mathbb{E}\left[\nabla^2_{\beta\beta}U_i \mid X_i\right]$$

and

$$V_{\nabla^2,i} = \nabla^2_{\beta\beta}\widehat{U}_i(t,d;\beta_0) - \mathbb{E}\left[\nabla^2_{\beta\beta}\widehat{U}_i(t,d;\beta_0) \mid X_i\right].$$

Lemma IX.3 *Under the condition of Proposition 4.1:*

1.

$$\sup_i \sup_{t,d} |B_{\nabla^2,i}| = O_{\mathbb{P}}(g^2);$$

2. For any $\alpha \in (0,1)$,

$$\sup_i \sup_{t,d} \left\{g^3 \left\|V_{\nabla^2,i}\right\|\right\} = O_{\mathbb{P}}(n^{-1/2}g^{\alpha/2});$$

3.

$$\sup_{t,d}\left|\frac{1}{n^2}\sum_{1\le i,j\le n} V_{\nabla^2,i}(t,d;\beta_0)U_j(t,d;\beta_0)\omega_{ij}\right| = o_{\mathbb{P}}(1);$$

4.

$$\sup_{t,d}\left|\frac{1}{n^2}\sum_{1\le i,j\le n} \mathbb{E}\left[\nabla^2_{\beta\beta}U_i(t,d;\beta_0) \mid X_i\right] U_j(t,d;\beta_0)\omega_{ij}\right| = o_{\mathbb{P}}(1).$$

Proof. 1. For a vector u, let $u^{\otimes 2}$ be the matrix $uu^\top$. For the uniform rate on the bias $B_{\nabla^2,i}$ let us write

$$\begin{aligned}
\mathbb{E}\left[\nabla^2_{\beta\beta}\widehat{U}_i(t,d;\beta_0) \mid X_i\right] &= H_{d,\beta_0}((-\infty,t] \mid X_i^\top\beta_0) \\
&\quad \times \mathbb{E}\left[\left(\widetilde{X}_i - \mathbb{E}[\widetilde{X}_k \mid X_k^\top\beta_0]\right)^{\otimes 2} g^{-3}L''((X_i - X_k)^\top\beta_0/g) \mid X_i\right] \\
&\quad - \mathbb{E}\left[H_{d,\beta_0}((-\infty,t] \mid X_k^\top\beta_0)\left(\widetilde{X}_i - \mathbb{E}[\widetilde{X}_k \mid X_k^\top\beta_0]\right)^{\otimes 2}\right. \\
&\quad \left. \times\, g^{-3}L''((X_i - X_k)^\top\beta_0/g) \mid X_i\right].
\end{aligned}$$

The uniform rate for $B_{\nabla^2,i}$ follows by integration by parts and Taylor expansion.

2. Let us simplify the notation and write $\mathcal{I}_i(t,d) = \mathbf{1}\{Y_i \le t; \delta_i = d\}$. To derive the uniform rate of $g^3 \|V_{\nabla^2,i}\|$ it suffices to consider the following sum of empirical processes

$$
\begin{aligned}
g^3 \left[\nabla^2_{\beta\beta}\widehat{U}_i(t,d;\beta_0) - \mathbb{E}\left[\nabla^2_{\beta\beta}\widehat{U}_i(t,d;\beta_0) \mid X_i\right]\right] \\
&= \frac{1}{n}\sum_{k=1}^{n}\Big\{\mathcal{I}_i(t,d)(\widetilde{X}_i - \widetilde{X}_k)^{\otimes 2}L''((X_i - X_k)^\top\beta_0/g) \\
&\quad -H_{d,\beta_0}((-\infty,t] \mid X_i^\top\beta_0)\mathbb{E}\left[(\widetilde{X}_i - \widetilde{X}_k)^{\otimes 2}L''((X_i - X_k)^\top\beta_0/g) \mid X_i\right]\Big\} \\
&\quad + \frac{1}{n}\sum_{k=1}^{n}\Big\{\mathcal{I}_k(t,d)(\widetilde{X}_i - \widetilde{X}_k)^{\otimes 2}L''((X_i - X_k)^\top\beta_0/g) \\
&\quad -\mathbb{E}\left[H_{d,\beta_0}((-\infty,t] \mid X_k^\top\beta_0)(\widetilde{X}_i - \widetilde{X}_k)^{\otimes 2}L''((X_i - X_k)^\top\beta_0/g) \mid X_i\right]\Big\}
\end{aligned}
$$

and to apply the Maximal Inequality and a change of variables.

3. Up to uniformly negligible diagonal terms,

$$
\frac{g^3}{n^2}\sum_{1\le i,j\le n} V_{\nabla^2,i}(t,d;\beta_0)U_j(t,d;\beta_0)\omega_{ij}
$$

is equal to the following sum of two centered $U-$processes of order three

$$
\begin{aligned}
&\frac{1}{n(n-1)(n-2)}\sum_{i\neq j\neq k}\Big\{\mathcal{I}_i(t,d)(\widetilde{X}_i - \widetilde{X}_k)^{\otimes 2}L''((X_i - X_k)^\top\beta_0/g) \\
&\quad -H_{d,\beta_0}((-\infty,t] \mid X_i^\top\beta_0)\mathbb{E}\left[(\widetilde{X}_i - \widetilde{X}_k)^{\otimes 2}L''((X_i - X_k)^\top\beta_0/g) \mid X_i\right]\Big\}U_j\omega_{ij} \\
&\quad + \frac{1}{n(n-1)(n-2)}\sum_{i\neq j\neq k}\Big\{\mathcal{I}_k(t,d)(\widetilde{X}_i - \widetilde{X}_k)^{\otimes 2}L''((X_i - X_k)^\top\beta_0/g) \\
&\quad -\mathbb{E}\left[H_{d,\beta_0}((-\infty,t] \mid X_k^\top\beta_0)(\widetilde{X}_i - \widetilde{X}_k)^{\otimes 2}L''((X_i - X_k)^\top\beta_0/g) \mid X_i\right]\Big\}U_j\omega_{ij}
\end{aligned}
$$

for which we consider the Hoeffding decomposition. It is easy to check that the sum of the two $U-$processes of order 1 in the decomposition vanishes. Moreover, among the conditional expectations given a couple of variables, only the conditional expectations given the couples of observations (k,j) do not vanish. By the Maximal Inequality and standard calculations, the sum of degenerate $U-$processes of order 3 (resp. order 2) in the decomposition has the uniform rate $O_{\mathbb{P}}(n^{-3/2}g^{\alpha/2})$ (resp. $O_{\mathbb{P}}(n^{-1}g^{\alpha/2})$), where $\alpha \in (0,1)$. Since α could be close to 1 and $ng^3 \to \infty$, the stated result follows.

4. The same arguments as used for the previous rate apply. ∎

Let us recall the decomposition

$$\begin{aligned}\frac{1}{2}\nabla^2_{\beta\beta}\widehat{I}_n(t,d;\beta) &= \frac{1}{n^2}\sum_{1\leq i,j\leq n}\nabla_\beta\widehat{U}_i(t,d;\beta)\nabla_\beta\widehat{U}_j(t,d;\beta)^\top\omega_{ij}\\ &+ \frac{1}{n^2}\sum_{1\leq i,j\leq n}\nabla^2_{\beta\beta}\widehat{U}_i(t,d;\beta)\widehat{U}_j(t,d;\beta)\omega_{ij}\\ &= E_{n1}(t,d;\beta)+E_{n2}(t,d;\beta).\end{aligned}$$

Let (b_n) be a sequence tending to zero such that

$$\mathbb{P}[\|\widehat{\beta}-\beta_0\|\leq b_n]\to 1.$$

Lemma IX.4 *Under the condition of Proposition 4.1,*

$$\sup_{t,d}\sup_{\|\beta-\beta_0\|\leq b_n}|E_{n1}(t,d;\beta)-E_{n1}(t,d;\beta_0)| = o_{\mathbb{P}}(1)$$

and

$$\sup_{t,d}\sup_{\|\beta-\beta_0\|\leq b_n}|E_{n2}(t,d;\beta)-E_{n2}(t,d;\beta_0)| = o_{\mathbb{P}}(1).$$

Proof. Let us simplify the notation and write $\mathcal{I}_i = \mathbf{1}\{Y_i\leq t;\delta_i=d\}$ and $H_i(\beta) = \mathbb{E}[\mathbf{1}\{Y_i\leq t;\delta_i=d\}\mid X_i^\top\beta]$. Let us decompose

$$E_{n1}(t,d;\beta) = E_{n11}(t,d;\beta)+E_{n12}(t,d;\beta),$$

where the (r,q) entry of the matrix $E_{n11}(t,d;\beta)$ is

$$E_{n11}(t,d;\beta)_{rq} = \frac{g^{-4}}{(n)_4}\sum_{\leq i\neq j\neq l\neq k}(\mathcal{I}_i-\mathcal{I}_k)(\mathcal{I}_j-\mathcal{I}_l)(\widetilde{X}_i-\widetilde{X}_k)_r(\widetilde{X}_j-\widetilde{X}_l)_q L'_{ik}(\beta,g)L'_{jl}(\beta,g)\omega_{ij},$$

and $E_{n1}(t,d;\beta)-E_{n11}(t,d;\beta)$. Here, $(n)_4 = n(n-1)(n-2)(n-3)$, $L'_{ik}(\beta,g) = L'((X_i-X_k)^\top\beta/g)$, L' is the derivative of L, and for a column vector V, $(V)_r$ denotes its rth component. We apply the Hoeffding decomposition to the fourth order $U-$processes $g^4\left[E_{n11}(t,d;\beta)_{rq}-E_{n11}(t,d;\beta_0)_{rq}\right]$. The degenerate $U-$processes of order four and three have the respective uniform rates $O_{\mathbb{P}}(n^{-2})$ and $O_{\mathbb{P}}(n^{-3/2})$. (These rates could be improved by a factor $g^{\alpha/2}$, but this is unnecessary at this stage.) Since $ng^3\to\infty$, these degenerate $U-$processes will be negligible. It remains to study the degenerate $U-$processes of order 2, 1 and the mean and bound their variations when β gets close to β_0. Let us remark that, since ω_{ij} does not depend on β and given some symmetry arguments, conditioning with the observations (i,j) is similar to do it with the variables (k,l), (i,l) is similar to (j,k) and (i,k) to (j,l). That means, only three types of conditional expectations given two observations, as involved in the degenerate $U-$process of order two, have to be investigated. Let us denote

$$\mathbb{E}_{ij}[\cdots] = \mathbb{E}[\cdots\mid Y_i,\delta_i,X_i,Y_j,\delta_j,X_j]\quad\text{and}\quad \mathbb{E}_i[\cdots] = \mathbb{E}[\cdots\mid Y_i,\delta_i,X_i]$$

and, for any given components r and q, simply denote

$$e_{ikjl}(\beta) = e_{ikjl}(\beta)_{rq} = (\mathcal{I}_i - \mathcal{I}_k)(\mathcal{I}_j - \mathcal{I}_l)(\widetilde{X}_i - \widetilde{X}_k)_r(\widetilde{X}_j - \widetilde{X}_l)_q L'_{ik}(\beta, g)L'_{jl}(\beta, g)\omega_{ij}.$$

Then, by elementary properties of the conditional expectation,

$$\begin{aligned}\mathbb{E}_{ij}[e_{ikjl}(\beta)] &= (\widetilde{X}_i)_r(\widetilde{X}_j)_q\omega_{ij}\mathbb{E}_i[(\mathcal{I}_i - H_k(\beta))L'_{ik}(\beta, g)]\mathbb{E}_j[(\mathcal{I}_j - H_l(\beta))L'_{jl}(\beta, g)] \\ &- (\widetilde{X}_i)_r\omega_{ij}\mathbb{E}_i[(\mathcal{I}_i - H_k(\beta))L'_{ik}(\beta, g)]\mathbb{E}_j[\mathbb{E}[(\widetilde{X}_l)_q \mid X_l^\top\beta](\mathcal{I}_j - H_l(\beta))L'_{jl}(\beta, g)] \\ &- (\widetilde{X}_j)_q\omega_{ij}\mathbb{E}_i[\mathbb{E}[(\widetilde{X}_k)_r \mid X_k^\top\beta](\mathcal{I}_i - H_k(\beta))L'_{ik}(\beta, g)]\mathbb{E}_j[(H_j(\beta) - H_l(\beta))L'_{jl}(\beta, g)] \\ &+ \omega_{ij}\mathbb{E}_i[\mathbb{E}[(\widetilde{X}_k)_r \mid X_k^\top\beta](\mathcal{I}_i - H_k(\beta))L'_{ik}(\beta, g)]\mathbb{E}_j[\mathbb{E}[(\widetilde{X}_l)_q \mid X_l^\top\beta](\mathcal{I}_j - H_l(\beta))L'_{jl}(\beta, g)].\end{aligned}$$

For each of the eight conditional expectations $\mathbb{E}_i$ or $\mathbb{E}_j$ on the right hand side of the last equation we could apply twice Lemma IX.5 to bound their variation when β gets close to β_0. For instance,

$$\begin{aligned}&\mathbb{E}_i[\mathbb{E}[(\widetilde{X}_k)_r \mid X_k^\top\beta](\mathcal{I}_i - H_k(\beta))g^{-2}L'_{ik}(\beta, g)] \\ &\qquad = \mathbb{E}_i[\mathbb{E}[(\widetilde{X}_k)_r \mid X_k^\top\beta_0](\mathcal{I}_i - H_k(\beta_0))g^{-2}L'_{ik}(\beta_0, g)] + R_i,\end{aligned}$$

where,

$$\sup_{1\leq r\leq p-1}\sup_{t,d}\sup_{\|\beta-\beta_0\|\leq b_n}|R_i| = \{1 + \|X_i\|\}o_{\mathbb{P}}(1),$$

and the $o_{\mathbb{P}}(1)$ factor does not depend on X_i. Moreover, the conditional expectation $\mathbb{E}_i[\mathbb{E}[(\widetilde{X}_k)_r \mid X_k^\top\beta_0](\mathcal{I}_i - H_k(\beta_0))g^{-2}L'_{ik}(\beta_0, g)]$ is uniformly bounded. Recall that, by our assumptions, $\sup_{1\leq i\leq n}\|\widetilde{X}_i\| = O_{\mathbb{P}}(\log n)$. Deduce that

$$\sup_{t,d}\sup_{\|\beta-\beta_0\|\leq b_n}|\mathbb{E}_{ij}[e_{ikjl}(\beta)] - \mathbb{E}_{ij}[e_{ikjl}(\beta_0)]|^2 = \{1 + \|X_i\|^2 + \|X_j\|^2\}g^8 o_{\mathbb{P}}(1) = g^8\log^2 n\ o_{\mathbb{P}}(1), \tag{IX.6}$$

and the $o_{\mathbb{P}}(1)$ factor does not depend on X_i or X_j. Moreover, the conditional expectation $g^{-4}\mathbb{E}_{ij}[e_{ikjl}(\beta)]/\log n$ is uniformly bounded. Thus the conditional expectations $g^{-4}\mathbb{E}_i[e_{ikjl}(\beta)]/\log n$, $g^{-4}\mathbb{E}_j[e_{ikjl}(\beta)]/\log n$ and the expectation $g^{-2}\mathbb{E}[e_{ikjl}(\beta)]/\log n$ are also uniformly bounded. Recall that $g^a\log n \to 0$ for any $a > 0$. Gathering facts, the contribution of $\mathbb{E}_{ij}[e_{ikjl}(\beta)] - \mathbb{E}_{ij}[e_{ikjl}(\beta_0)]$ to the right hand side of the maximal inequality, applied with the degenerate $U-$process of order two in the Hoeffding decomposition of $g^4\left[E_{n11}(t, d; \beta)_{rq} - E_{n11}(t, d; \beta_0)_{rq}\right]$, is of rate $O(g^{4\alpha}\log^\alpha n)$ with $\alpha \in (0, 1)$ that could be arbitrarily close to 1. Finally, this will produce a contribution of uniform rate $O_{\mathbb{P}}(n^{-1}g^{4\alpha-4}\log^\alpha n) = o_{\mathbb{P}}(1)$ for the $U-$processes of order two.

Next, let us investigate the variations of $\mathbb{E}_{il}[e_{ikjl}(\beta)]$. In this case, one still decompose $\mathbb{E}_{il}[e_{ikjl}(\beta)]$ in a sum of terms, each of them under the form of products involving conditional expectations given only one observation. Then, the same uniform rate as in equation (IX.6) could be derived. Moreover, by similar arguments, the conditional expectation $g^{-4}\mathbb{E}_l[e_{ikjl}(\beta)]/\log n$ and the expectation $g^{-4}\mathbb{E}[e_{ikjl}(\beta)]/\log n$ could be shown to be uniformly bounded. One could conclude that $\mathbb{E}_{il}[e_{ikjl}(\beta)] - \mathbb{E}_{il}[e_{ikjl}(\beta_0)]$ will behave similarly to $\mathbb{E}_{ij}[e_{ikjl}(\beta)] - \mathbb{E}_{ij}[e_{ikjl}(\beta_0)]$ uniformly in a neighborhood of β_0.

To complete the analysis of the degenerate $U-$processes of order 2, it remains to investigate the variations of $\mathbb{E}_{ik}[e_{ikjl}(\beta)]$. We have

$$
\begin{aligned}
g^{-2}\mathbb{E}_{ik}[e_{ikjl}(\beta)_{rq}] &= (\mathcal{I}_i - \mathcal{I}_k)(\widetilde{X}_i - \widetilde{X}_k)_r L'_{ik}(\beta, g)\mathbb{E}_i[(\mathcal{I}_j - \mathcal{I}_l)(\widetilde{X}_j - \widetilde{X}_l)_q g^{-2} L'_{jl}(\beta, g)\omega_{ij}] \\
&= (\mathcal{I}_i - \mathcal{I}_k)(\widetilde{X}_i - \widetilde{X}_k)_r L'_{ik}(\beta, g)\mathbb{E}_i[\mathbb{E}[(\widetilde{X}_j)_q \mathcal{I}_j \omega_{ij} \mid X_i, X_j^\top\beta]\mathbb{E}[g^{-2}L'_{jl}(\beta, g) \mid X_i, X_j^\top\beta]] \\
&- (\mathcal{I}_i - \mathcal{I}_k)(\widetilde{X}_i - \widetilde{X}_k)_r L'_{ik}(\beta, g)\mathbb{E}_i[\mathbb{E}[\mathcal{I}_j \omega_{ij} \mid X_i, X_j^\top\beta]\mathbb{E}_l[(\widetilde{X}_l)_q \mid X_i, X_l^\top\beta] g^{-2} L'_{jl}(\beta, g)] \\
&- (\mathcal{I}_i - \mathcal{I}_k)(\widetilde{X}_i - \widetilde{X}_k)_r L'_{ik}(\beta, g)\mathbb{E}_i[\mathbb{E}[\mathcal{I}_l \mid X_i, X_l^\top\beta]\mathbb{E}_l[(\widetilde{X}_j)_q \omega_{ij} \mid X_i, X_j^\top\beta] g^{-2} L'_{jl}(\beta, g)] \\
&+ (\mathcal{I}_i - \mathcal{I}_k)(\widetilde{X}_i - \widetilde{X}_k)_r L'_{ik}(\beta, g)\mathbb{E}_i[\mathbb{E}[(\widetilde{X}_l)_q \mathcal{I}_l \omega_{ij} \mid X_i, X_l^\top\beta]\mathbb{E}[g^{-2}L'_{jl}(\beta, g) \mid X_i, X_l^\top\beta]].
\end{aligned}
$$

Using the fact that $\sup_{1\leq i\leq n}\|X_i\| = O_{\mathbb{P}}(\log n)$, by Lemma IX.5,

$$
\mathbb{E}[g^{-2}L'_{jl}(\beta, g) \mid X_i, X_j^\top\beta] = \mathbb{E}[g^{-2}L'_{jl}(\beta_0, g) \mid X_j^\top\beta_0] + o_{\mathbb{P}}(\log n),
$$

uniformly with respect to $\|\beta - \beta_0\| \leq b_n$. On the other hand, using arguments as in Lemma IX.5, for some suitable functions γ_β,

$$
\mathbb{E}[\gamma_\beta(X_i^\top\beta, X_k^\top\beta; t, d)g^{-2}[L'_{ik}(\beta, g)]^2] = \mathbb{E}[\gamma_\beta(X_i^\top\beta_0, X_k^\top\beta_0; t, d)g^{-2}[L'_{ik}(\beta, g)]^2] + o_{\mathbb{P}}(\log n),
$$

uniformly. Moreover, for uniformly bounded functions γ_β, since $L'(\cdot)$ is bounded and integrable, $\mathbb{E}[\gamma_\beta(X_i^\top\beta_0, X_k^\top\beta_0; t, d)g^{-2}[L'_{ik}(\beta, g)]^2]$ is uniformly bounded. Repeating the arguments several times, deduce that

$$
\mathbb{E}\left[\sup_{t,d}\sup_{\|\beta-\beta_0\|\leq b_n}\{\mathbb{E}_{ik}[e_{ikjl}(\beta)_{rq}] - \mathbb{E}_{ik}[e_{ikjl}(\beta_0)_{rq}]\}^2\right] = g^6 O(\log^4 n),
$$

and thus, by the Maximal Inequality, $\mathbb{E}_{ik}[e_{ikjl}(\beta)] - \mathbb{E}_{ik}[e_{ikjl}(\beta_0)]$ is uniformly negligible in a neighborhood of β_0. The degenerate $U-$processes of order 1 and the mean could be handled with similar arguments. For the mean, one slight difference is that instead of taking the supremum of quantities like $\|X_i\|$ that appears in the bound given by Lemma IX.5, one has to integrate them out. This avoids to pay the price of a $\log n$ factor in the bounds and hence yields $\mathbb{E}[e_{ikjl}(\beta)] - \mathbb{E}[e_{ikjl}(\beta_0)] = o_{\mathbb{P}}(1)$ uniformly with respect to t and d, and uniformly over $o_{\mathbb{P}}(1)$ neighborhoods of β_0.

Up to a factor that converges to 1, the quantity $E_{n12}(t, d; \beta)$ contains the diagonal terms of $E_{n1}(t, d; \beta)$ and thus is negligible. The arguments for $E_{n2}(t, d; \beta)$ are very similar and hence we omit the details. Let us only mention that for the degenerate $U-$process of order 1 and for the mean, we use the fact that, by our assumptions and for suitable functions γ_β, quantities like $\mathbb{E}_i\left[\gamma_\beta(X_i^\top\beta, X_k^\top\beta; t, d)g^{-3}L''_{ik}(\beta, g)\right]$ are uniformly bounded. ■

Lemma IX.5 *Let $L^{(\kappa)}(\cdot)$ denote the κth derivative of $L(\cdot)$, $\kappa \in \{0,1,2\}$. Let b_n, $n \geq 1$ be a positive sequence of real numbers converging to zero. Let $z \mapsto \lambda_\beta(z;t,d)$, $z \in \mathbb{R}$ be a class of κ times differentiable functions indexed by β, t and d such that*

$$\sup_{\|\beta-\beta_0\|\leq b_n} \sup_{t,d} \sup_{z\in\mathbb{R}} \left| \partial_z^{(\kappa)}[\lambda_\beta(\cdot;t,d)f_\beta(\cdot)](z) - \partial_z^{(\kappa)}[\lambda_{\beta_0}(\cdot;t,d)f_{\beta_0}(\cdot)](z) \right| = 0,$$

and $\partial_z^{(\kappa)}[\lambda_{\beta_0}(\cdot;t,d)f_{\beta_0}(\cdot)](\cdot)$ is uniformly bounded, where $\partial_z^{(\kappa)}$ denotes the derivative of order κ with respect to z for $\kappa \in \{0,1,2\}$. If the function $z \mapsto \partial_z^{(\kappa)}[\lambda_{\beta_0}(\cdot;t,d)f_{\beta_0}(\cdot)](z)$ is Lipschitz continuous with constant C independent of t and d, then

$$\begin{aligned}\sup_{t,d} \sup_{\|\beta-\beta_0\|\leq b_n} \Big| &\mathbb{E}\left[\lambda_\beta(X^\top\beta;t,d)g^{-\kappa-1}L^{(\kappa)}((x-X)^\top\beta/g)\right] \\ &-\mathbb{E}\left[\lambda_{\beta_0}(X^\top\beta_0;t,d)g^{-\kappa-1}L^{(\kappa)}((x-X)^\top\beta_0/g)\right]\Big| = \{1+\|x\|\}o_{\mathbb{P}}(1), \quad \text{(IX.7)}\end{aligned}$$

and the $o_{\mathbb{P}}(1)$ factor does not depend on x.

Proof of Lemma IX.5. Let simplify notation and write $\lambda_\beta(\cdot)$ instead of $\lambda_\beta(\cdot;t,d)$. First we consider the case $\kappa = 0$. By a standard change of variables and the stated assumptions, uniformly with respect to t, d, $\|\beta - \beta_0\| \leq b_n$, for any $x \in \mathcal{X}$,

$$\begin{aligned}\mathbb{E}\left[\lambda_\beta(X^\top\beta)g^{-1}L((x-X)^\top\beta/g)\right] &= \int_{\mathbb{R}} \lambda_\beta(z)g^{-1}L((x^\top\beta - z)/g)f_\beta(z)dz \\ &= \int_{\mathbb{R}} (\lambda_\beta f_\beta)(x^\top\beta - uh)L(u)du \\ &= \int_{\mathbb{R}} (\lambda_{\beta_0} f_{\beta_0})(x^\top\beta - uh)L(u)du + o(1) \\ &= \int_{\mathbb{R}} (\lambda_{\beta_0} f_{\beta_0})(x^\top\beta_0 - uh)L(u)du + O(b_n) + \|x\|O(b_n) \\ &= \mathbb{E}\left[\lambda_{\beta_0}(X^\top\beta_0)g^{-1}L((x-X)^\top\beta_0/g)\right] + \{1+\|x\|\}o(1).\end{aligned}$$

When $\kappa = 1$, before applying the uniform continuity condition (IX.7) and the Lipschitz property, we use integration by parts to deduce

$$\int_{\mathbb{R}} \lambda_\beta(z)f_\beta(z)g^{-2}L'((x^\top\beta - z)/g)dz = \int_{\mathbb{R}} \partial_z[\lambda_\beta f_\beta](z)g^{-1}L((x^\top\beta - z)/g)dz.$$

Next we can continue as in the case $\kappa = 0$. The case $\kappa = 2$ is similar and hence we omit the details. ∎

Lemma IX.6 *Under the Assumptions of Proposition 5.1, for any $b_n = O_{\mathbb{P}}(n^{-1/2})$,*

$$\sup_{\|\beta-\beta_0\|\leq b_n} \sup_{t\leq\tau} \sup_{x\in\overline{\mathcal{X}}} |\Delta_{1n}(t,x;\beta) - \Delta_{1n}(t,x;\beta_0)| = o_{\mathbb{P}}(n^{-1/2}h^{-1/2})$$

and

$$\sup_{\|\beta-\beta_0\|\leq b_n} \sup_{t\leq\tau} \sup_{x\in\overline{\mathcal{X}}} |\Delta_{2n}(t,x;\beta,\beta_0)| = o_{\mathbb{P}}(n^{-1/2}h^{-1/2}).$$

Proof of Lemma IX.6. First, we investigate the uniform rate of Δ_{2n}. Let

$$\xi(Y,\delta;t) = \mathbf{1}\{Y \le t\}\delta \quad \text{or} \quad \xi(Y,\delta;t) = \mathbf{1}\{Y \ge t\}, \qquad t \le \tau,$$

and let $\widetilde{H}_\beta(t \mid z)$ be any of $H_{1,\beta}((-\infty,t] \mid z) = \mathbb{E}\left[\mathbf{1}\{Y \le t\}\delta \mid X^\top\beta = z\right]$ or

$$H_\beta([t,\infty) \mid z) = \mathbb{E}\left[\mathbf{1}\{Y \ge t\} \mid X^\top\beta = z\right], \qquad t \le \tau.$$

By standard changes of variables,

$$\begin{aligned}
&\mathbb{E}\left[h^{-1}\xi(Y,\delta;t)\left\{K((x-X)^\top\beta/h) - K((x-X)^\top\beta_0/h)\right\}\right] \\
&\quad = \mathbb{E}\left[h^{-1}\widetilde{H}_\beta(t \mid X^\top\beta)K((x-X)^\top\beta/h) - \widetilde{H}_{\beta_0}(t \mid X^\top\beta_0)K((x-X)^\top\beta_0/h)\right] \\
&= \int_{\mathbb{R}} K(u)\left[\widetilde{H}_\beta(t \mid x^\top\beta - uh)f_\beta(x^\top\beta - uh) - \widetilde{H}_{\beta_0}(t \mid x^\top\beta - uh)f_{\beta_0}(x^\top\beta - uh)\right] du \\
&+ \int_{\mathbb{R}} K(u)\left[\widetilde{H}_{\beta_0}(t \mid x^\top\beta - uh)f_{\beta_0}(x^\top\beta - uh) - \widetilde{H}_{\beta_0}(t \mid x^\top\beta_0 - uh)f_{\beta_0}(x^\top\beta_0 - uh)\right] du.
\end{aligned}$$

Since $\|\beta - \beta_0\| \le b_n$, by the Lipschitz property of $\widetilde{H}_{\beta_0}(t \mid \cdot)f_{\beta_0}(\cdot)$, one could deduce

$$\sup_{\|\beta-\beta_0\|\le b_n}\sup_{t\le\tau}\sup_{x\in\overline{\mathcal{X}}}\left|\mathbb{E}\left[h^{-1}\xi(Y,\delta;t)\left\{K((x-X)^\top\beta/h) - K((x-X)^\top\beta_0/h)\right\}\right]\right| \lesssim b_n.$$

For the uniform rate for Δ_{1n}, let it suffice to consider the empirical process defined by the family of functions

$$\mathcal{F}_n = \{\xi(\cdot,\cdot;t)\left[K((\cdot-x)^\top\beta/h) - K((\cdot-x)^\top\beta_0/h)\right] : t \le \tau, x \in \overline{\mathcal{X}}, \|\beta-\beta_0\| \le b_n\}.$$

By the result of Nolan & Pollard (1987), Pakes and Pollard (1989), see also Sherman (1994) and van der Vaart & Wellner (1996), this class is a VC-class for a constant envelope, with constants A and V independent of n. By the Maximal Inequality applied for some α close to 1, the fact that $K(\cdot)$ is bounded, the uniform bound on the expectation of the function from $\mathcal{F}_n$ derived above, one could deduce that

$$\sup_{\|\beta-\beta_0\|\le b_n}\sup_{t\le\tau}\sup_{x\in\overline{\mathcal{X}}}|\Delta_{1n}(t,x;\beta) - \Delta_{1n}(t,x;\beta_0)| = O_{\mathbb{P}}(n^{-1/2}h^{-1}b_n^{\alpha/2}) = o_{\mathbb{P}}(n^{-1/2}h^{-1/2}).$$

■

Lemma IX.7 *Under the condition of Proposition 4.1*

$$\int \mathcal{V}_{1n}(t,d;\beta_0)dF_{n,Y,\delta}(t,d) = \int \mathcal{V}_{1n}(t,d;\beta_0)dF_{Y,\delta}(t,d) + o_{\mathbb{P}}(n^{-1/2}),$$

Proof. Let

$$v_{1j}(t,d;\beta_0)=\mathbb{E}\left\{\mathbb{E}\left[\nabla_\beta U(t,d;\beta_0)\mid X\right]\omega(X-X_j)\mid X_j\right\}U_j(t,d;\beta_0)$$

so that

$$\mathcal{V}_{1n}(t,d;\beta_0)=\frac{1}{n}\sum_{1\le j\le n}v_{1j}(t,d;\beta_0).$$

Then,

$$\begin{aligned}\int\mathcal{V}_{1n}(t,d;\beta_0)dF_{n,Y,\delta}(t,d)&=\frac{n-1}{n}\frac{1}{n(n-1)}\sum_{1\le i\ne j\le n}v_{1j}(Y_i,\delta_i;\beta_0)+\frac{1}{n^2}\sum_{1\le j\le n}v_{1j}(Y_j,\delta_j;\beta_0)\\&=\frac{n-1}{n}V_{n11}+V_{n12}.\end{aligned}$$

By the law of large numbers, it is easy to see that $V_{n12}=O_{\mathbb{P}}(n^{-1})$. On the other hand,

$$\begin{aligned}&\int v_{1j}(t,d;\beta_0)dF_{Y,\delta}(t,d)\\&\quad=\mathbb{E}\left[\mathbb{E}\left\{\mathbb{E}\left[\nabla_\beta U(\tilde{Y},\tilde{\delta};\beta_0)\mid X\right]\omega(X-X_j)\mid X_j,\tilde{Y},\tilde{\delta}\right\}U_j(\tilde{Y},\tilde{\delta};\beta_0)\mid Y_j,\delta_j,X_j\right].\end{aligned}$$

Next note that, since

$$\begin{aligned}&\mathbb{E}\left[v_{1j}(Y_i,\delta_i;\beta_0)\mid Y_i,\delta_i,X_i\right]\\&\quad=\mathbb{E}\left[\mathbb{E}\left\{\mathbb{E}\left[\nabla_\beta U(\mid Y_i,\delta_i;\beta_0)\mid X,Y_i,\delta_i\right]\omega(X-X_j)\mid X_j\right\}\mathbb{E}[U_j(t,d;\beta_0)\mid X_j]\mid Y_i,\delta_i,X_i\right]\\&\quad=0\end{aligned}$$

and

$$\mathbb{E}\left[\int v_{1j}(t,d;\beta_0)dF_{Y,\delta}(t,d)\mid Y_i,\delta_i,X_i\right]=0,$$

we have

$$\mathbb{E}\left[v_{1j}(Y_i,\delta_i;\beta_0)-\int v_{1j}(t,d;\beta_0)dF_{Y,\delta}(t,d)\mid Y_i,\delta_i,X_i\right]=0$$

Moreover, by construction,

$$\mathbb{E}\left[v_{1j}(Y_i,\delta_i;\beta_0)-\int v_{1j}(t,d;\beta_0)dF_{Y,\delta}(t,d)\mid Y_j,\delta_j,X_j\right]=0.$$

This means that $V_{n11}-\int\mathcal{V}_{1n}(t,d;\beta_0)dF_{Y,\delta}(t,d)$ is a degenerate $U-$statistics of order 2 and thus, by a simple variance calculation,

$$V_{n11}=\int\mathcal{V}_{1n}(t,d;\beta_0)dF_{Y,\delta}(t,d)+O_{\mathbb{P}}(n^{-1}).$$

Gathering the rates, the statement follows. ■

Additional References

1. Nolan, D., and Pollard, D. (1987). $U-$processes : Rates of convergence. *Annals of Statistics* **15**, 780–799.

2. Pakes, A., and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57**, 1027–1057.

3. van der Vaart, A.D. & Wellner, J.A. (1996). *Weak convergence and empirical processes.* Springer Series in Statistics. Springer-Verlag, New-York.